Karlin, Michael

From: Percival, Aaron (IC) <aaron.percival@canada.ca>
Sent: May 23, 2017 3:11 PM
To: Karlin, Michael
Subject: RE: AI and digital services

Hi, Michael.

Thanks for getting in touch.

I'm currently starting a program on client experience at the Canadian Intellectual Property Office (CIPO). My background, however, was as a physicist for Defence Research and Development Canada (DRDC) where we were using AI and machine learning for various applications.

I'm pushing CIPO to start investigating areas where machine learning can have a significant positive impact in client experience and process efficiency. Some of the obvious are more on the robotic process automation (RPA) side of the spectrum here, but there are ways to apply more sophisticated techniques.

Happy to discuss. How would you like to proceed?

Aaron

Aaron M. Percival, M.Sc.

Free Agent – IN·spire, NRCan's Innovation Hub
Posted to,

Innovation Officer, CIPO Business Services Branch
Innovation, Science and Economic Development Canada / Government of Canada
Aaron.Percival@canada.ca / Tel: 613-240-3729

Agent libre – IN·spire, Carrefour d'Innovation de RNCan
Affecté à,

Agent d'innovation, OPIC Direction des services aux entreprises
Innovation, Sciences et Développement économique Canada / Gouvernement du Canada
Aaron.Percival@canada.ca / Tél: 613-240-3729

From: Karlin, Michael
Sent: May-19-17 10:21 AM
To: Percival, Aaron (IC)
Subject: AI and digital services

Hi Aaron,

The PCO comms team sent me your way. I hear you're interested in AI and machine learning!

I'm leading some early work at TBS on a policy and ethical framework for AI applications in the area of digital services. We are seeing keen interest in service departments for using AI tools such as virtual agents or automated eligibility decision making to improve service delivery to the public. I want to leave the details of business cases to line

departments, but it is likely that they will face similar ethical, policy, and legal questions, so some central coordination in this area might be useful.

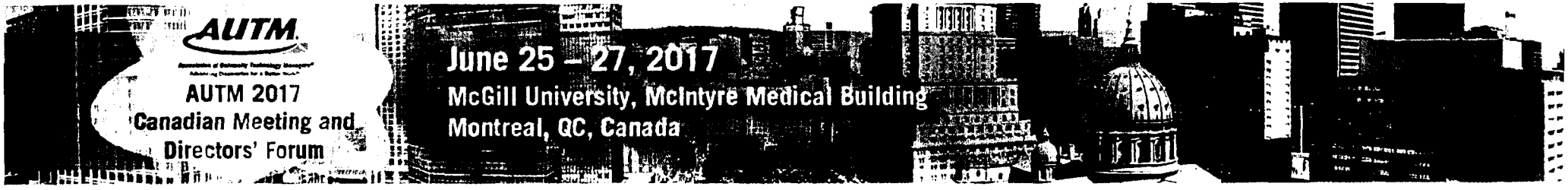
Is your team currently looking at AI functionality for your business? If so, I'd be really interested in discussing with you further.

Cheers,
M.

Michael Karlin

Advisor/Economist, Chief Information Officer Branch
Treasury Board of Canada Secretariat / Government of Canada
Michael.Karlin@tbs-sct.gc.ca

Conseiller/Économiste, Direction du dirigeant principal de l'information
Secrétariat du Conseil du Trésor du Canada / Gouvernement du Canada
Michael.Karlin@tbs-sct.gc.ca



Artificial Intelligence – Emerging Issues

Maya Medeiros

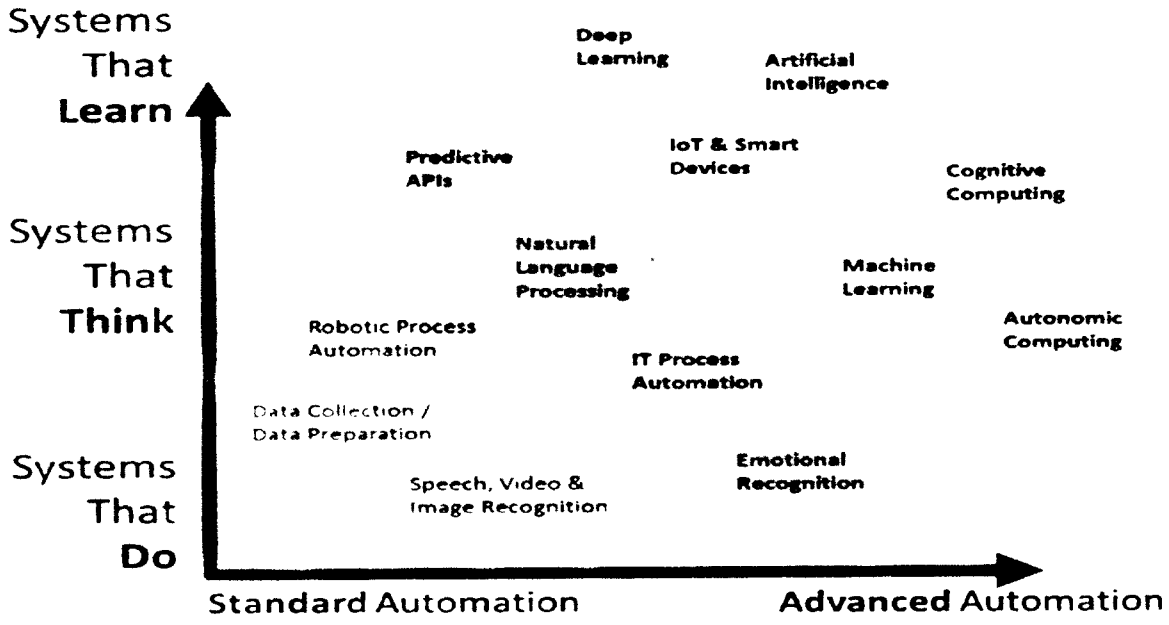
Partner, Norton Rose Fulbright LLP

- ***Success in creating AI could be the biggest event in the history of our civilisation. But it could also be the last, unless we learn how to avoid the risks. Alongside the benefits, AI will also bring dangers, like powerful autonomous weapons, or new ways for the few to oppress the many. It will bring great disruption to our economy.***

- Professor Stephen Hawking, quoted in Maya Oppenheim, *Stephen Hawking: Artificial Intelligence Could be the Greatest Disaster in Human History*, The Independent, 20 October 2016

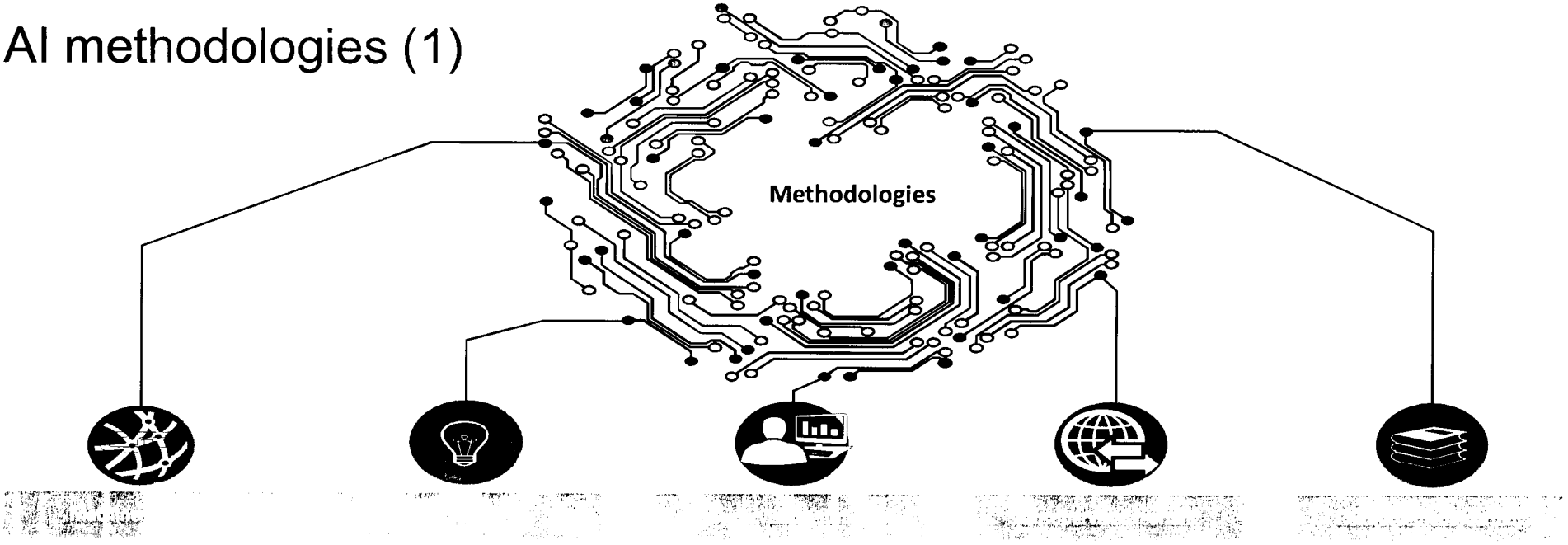
- ***AI is a field of computer science that includes machine learning, natural language processing, speech processing, expert systems, robotics, and machine vision.***

AI encompasses a wide spectrum of technologies



Source: Bart Van der Mark, A Primer On Robotic Process Automation, digitally.cognizant.com, 27 January 2016

AI methodologies (1)



- Automating decision-making using programming rules and training data sets
- Human subject-matter experts provide feedback on results as part of a training process
- Can adapt its programming based on the training and feedback

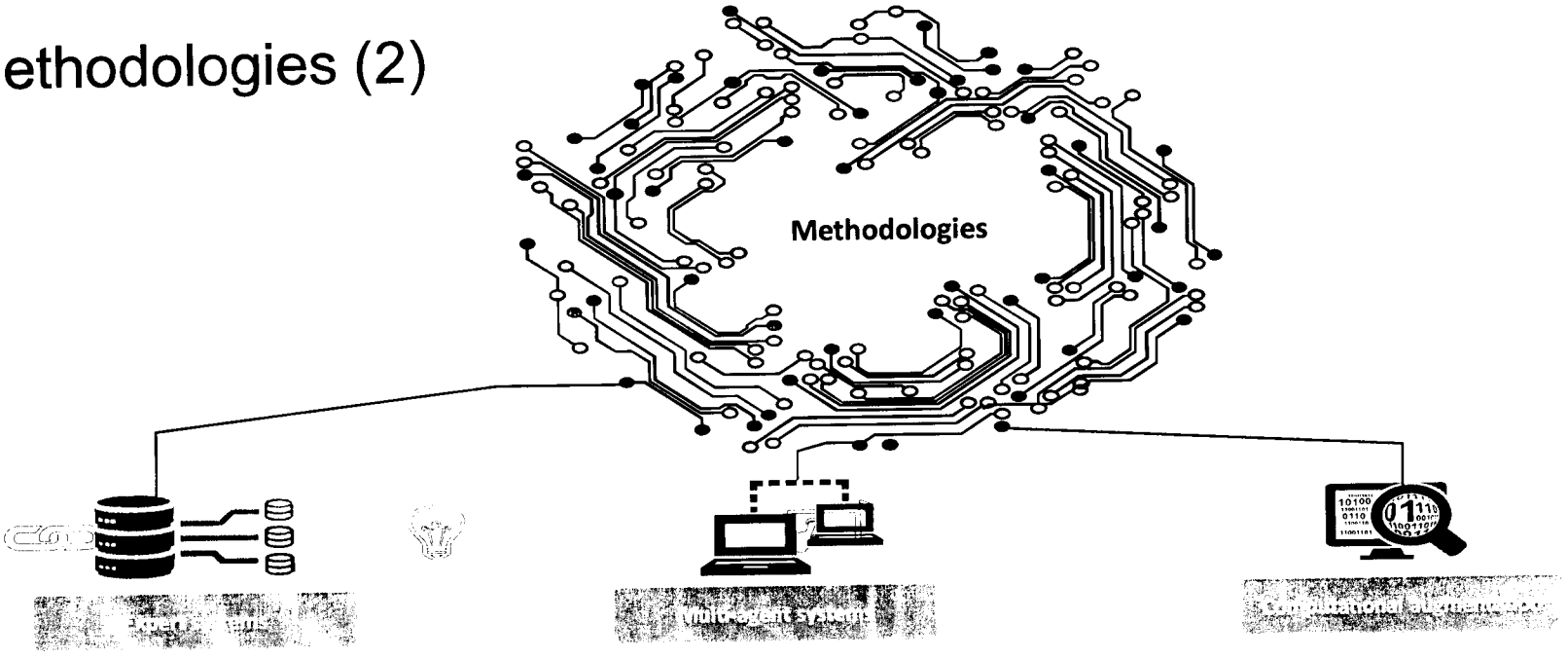
- Use of multiple layers of abstract representations of data to optimise the machine learning process

- Labelled training data examples are used to infer functions that can be used for processing new data
- Computer can predict or "guess at" the meaning of new data based on the training data set, graph and network structures, and feedback

- Labelling data based on inferences about its structure

- Rules to control software action in an environment to maximise a reward
- May not need training data examples with labelled data sets

AI methodologies (2)



- Inductive reasoning based mainly on "if–then" rules or logic programming

- Machine learning in combination with multiple intelligent entities (for example, so-called "swarms")

- Methodology addressing computational challenges characterised by lack of certain, consistent and complete information, and when numerical (e.g. statistical) information is only partially / not available

AI applications include …

Speech processing

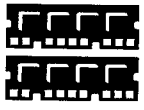
– Conversion between speech (audio) and text

Natural language processing

– Deriving meaning, context, or sentiment in textual data or conversations with humans using grammars and graph structures

Machine vision

– Detecting patterns in visual content for object tracking, audio, and face recognition



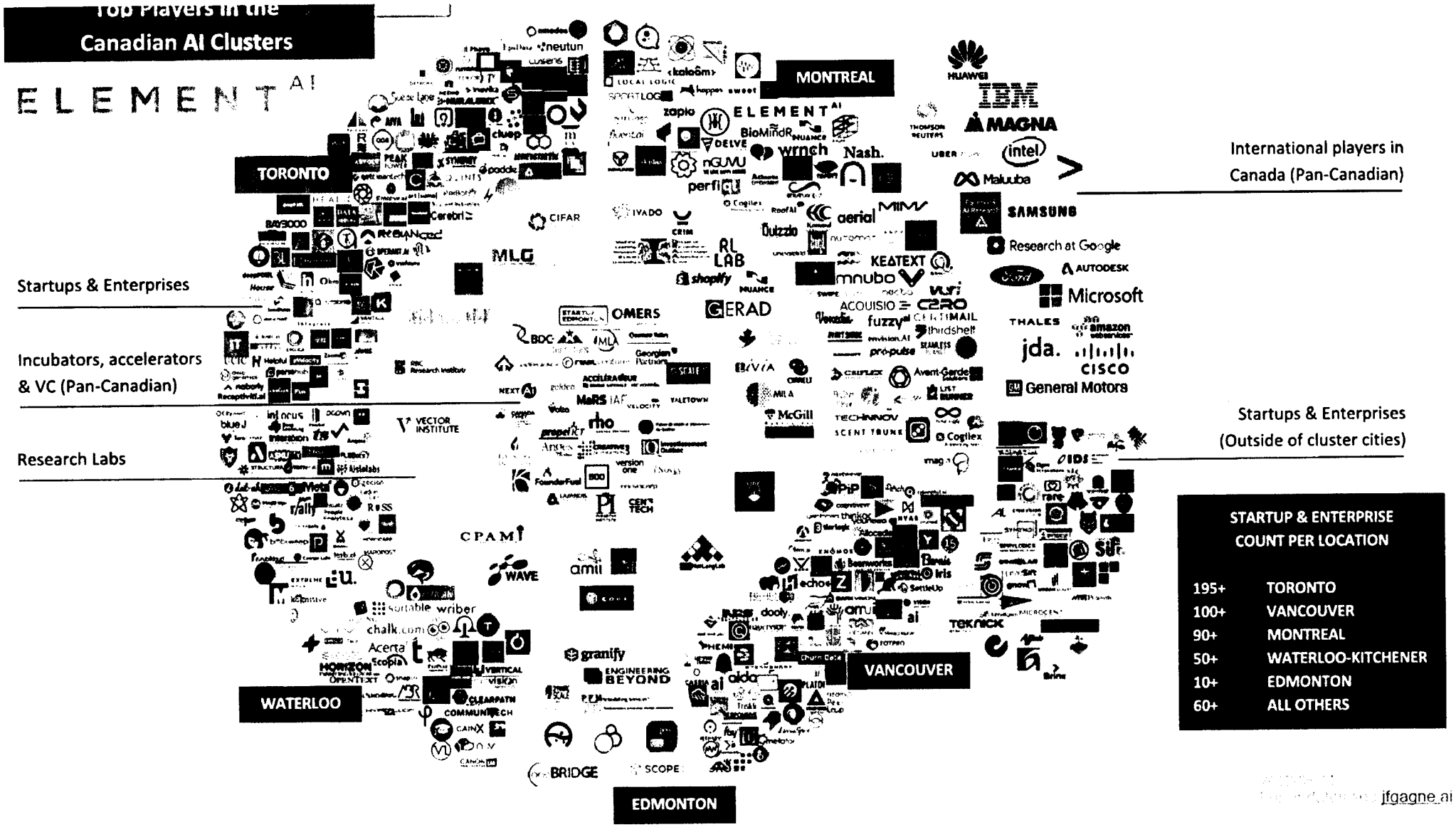
Robotics

– the use of AI systems to automate and mechanically control machine movements

AI planning

– A form of automated programming

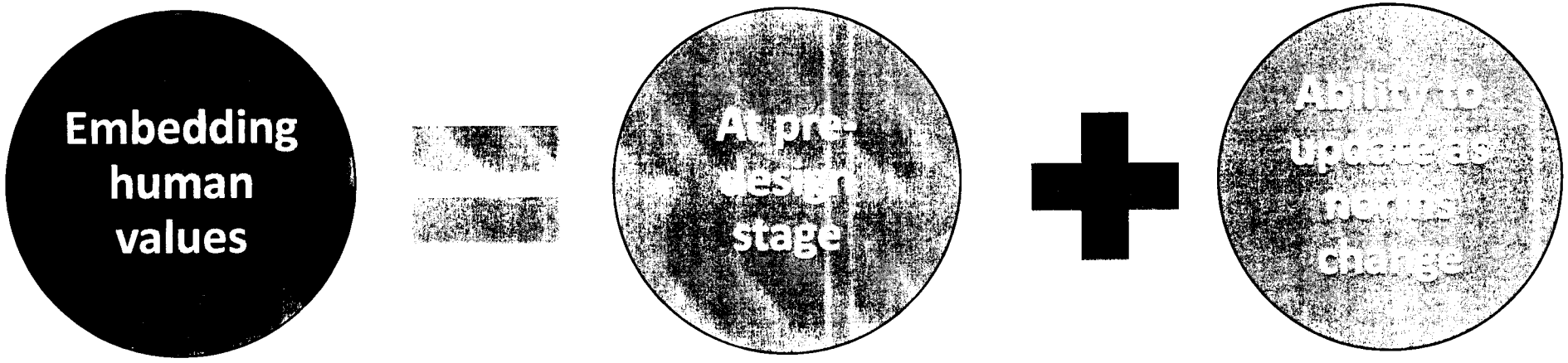
Top Players in the Canadian AI Clusters



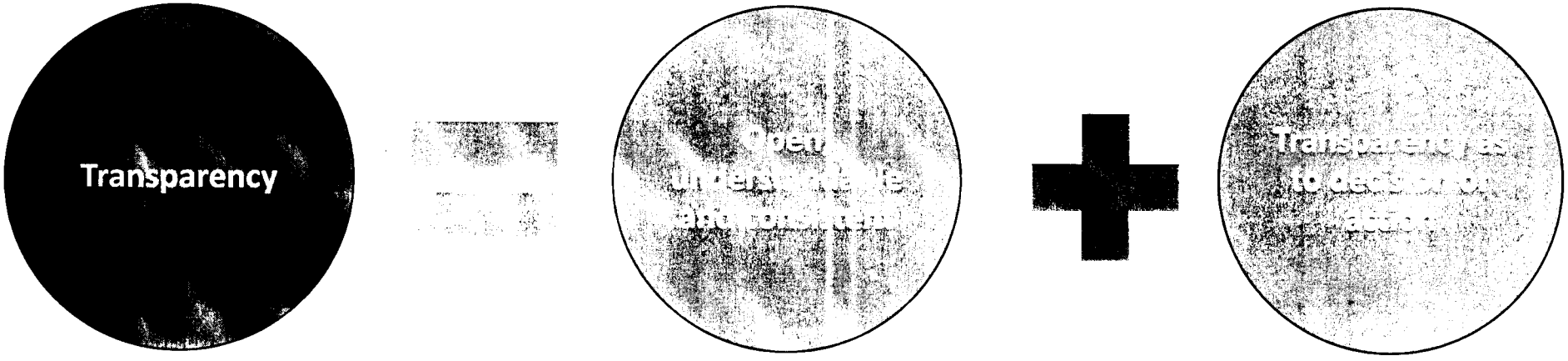
jfgagne.ai

- *For AI to be accepted for use in a given market (for example, by achieving sufficient end user uptake), as a matter of commercial reality the use of AI will need to be perceived by the participants in that market as meeting certain minimum ethical standards. What these are will vary according to the type of AI at issue and the relevant sector in which it is deployed.*

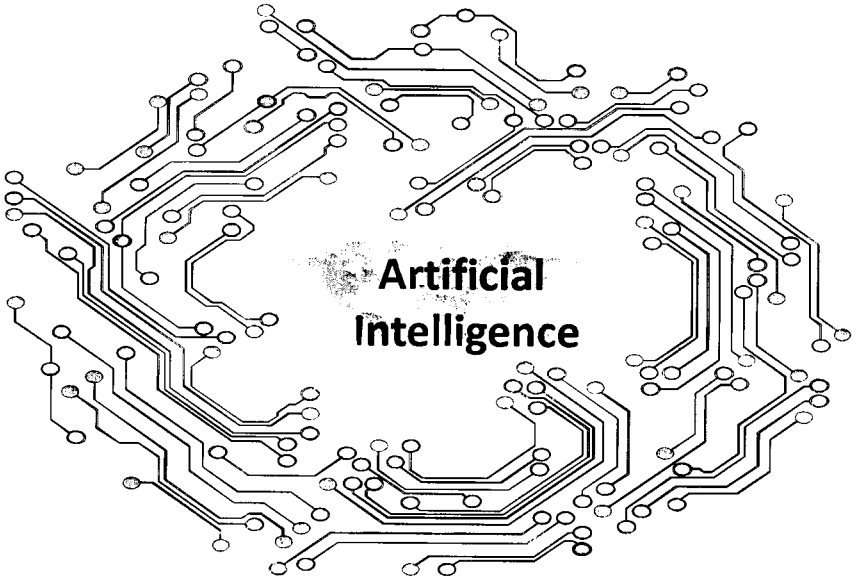
Ethical AI requires … (1)



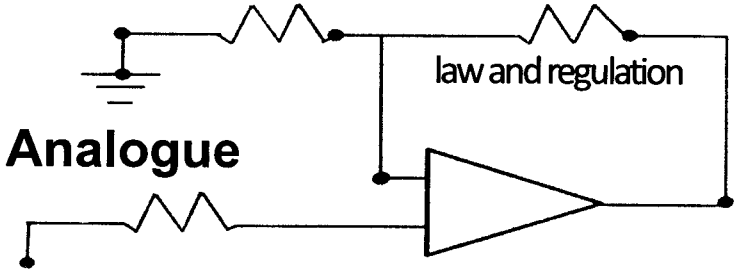
Ethical AI requires … (2)



Ethical AI requires … (3)

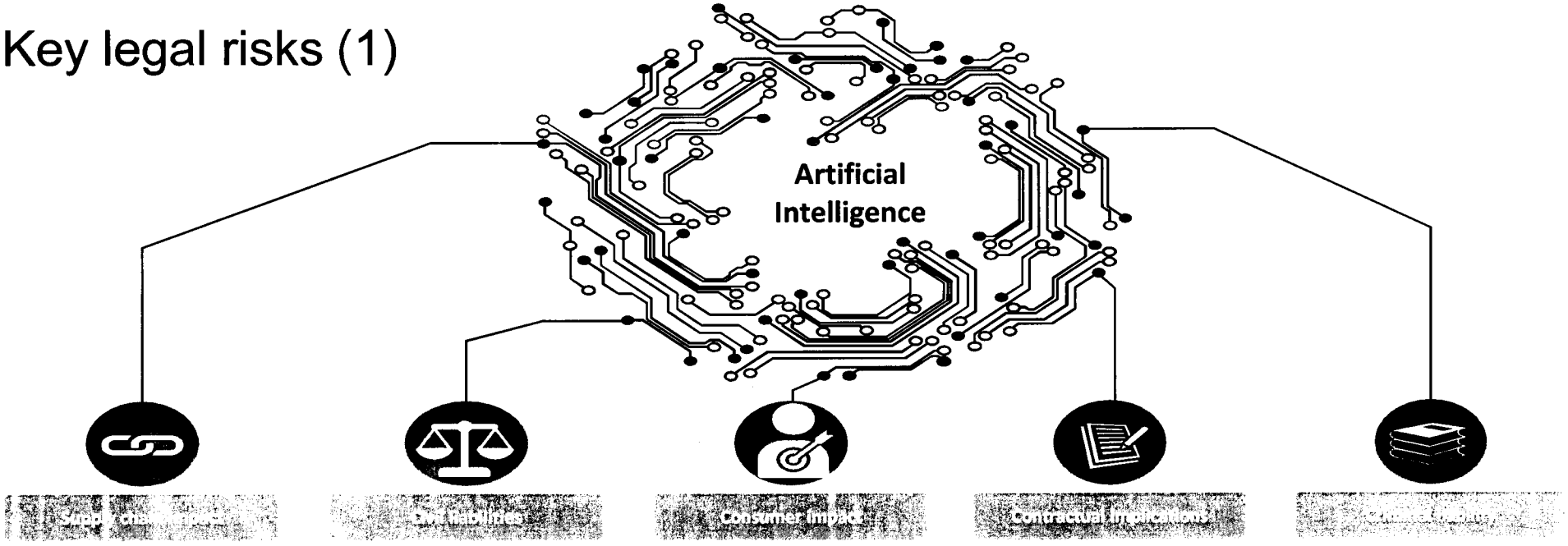


V



We are entering into an AI digital era governed by analogue law and regulation

Key legal risks (1)



Does the development / use of AI by the business change which AI supply chain participant(s) may be potentially liable in connection with it, compared with how liability is typically allocated now?

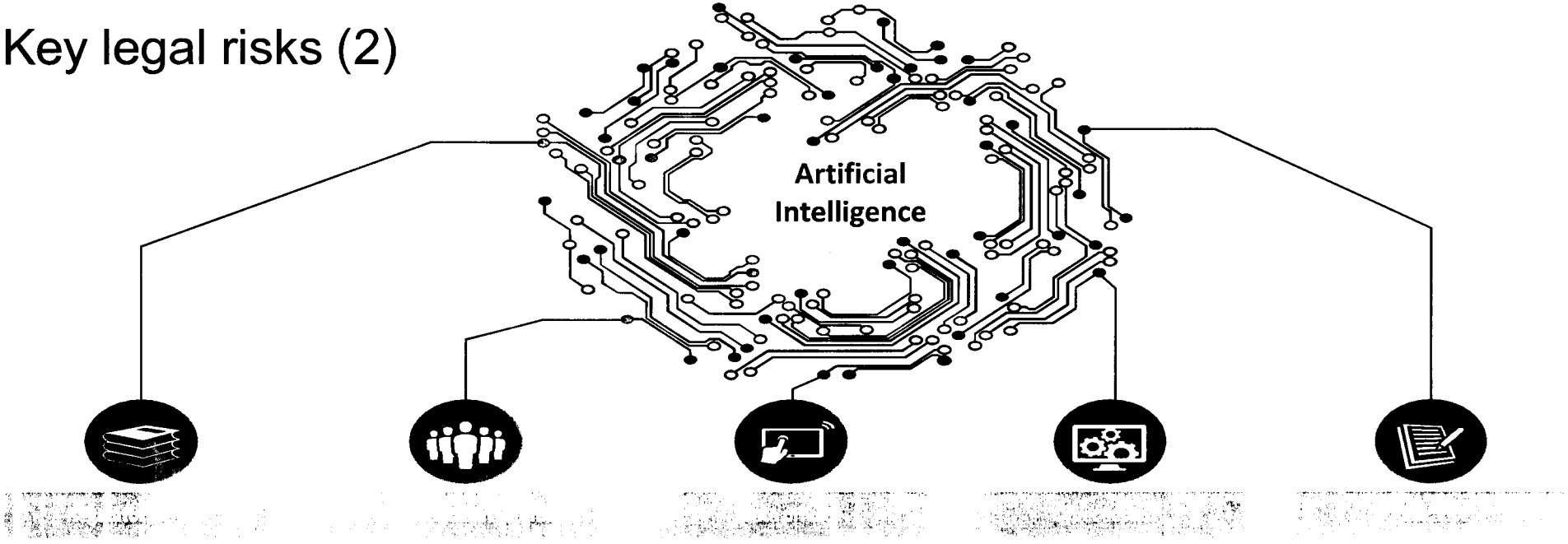
What types of civil liabilities might arise for the business, or in connection with its supply chain?

If consumers could be affected by AI developed or used by the business, could the use of AI in relation to consumers adversely affect the liability profile of the business?

Might any such liabilities need to be reallocated by new contractual liability or indemnification schemes where appropriate?

Is there a risk of criminal liability for the business in connection with AI?

Key legal risks (2)



Will any regulations apply that could be relevant in relation to a business's development or use of AI (including environmental and health and safety regulations)? Are new regulations likely?

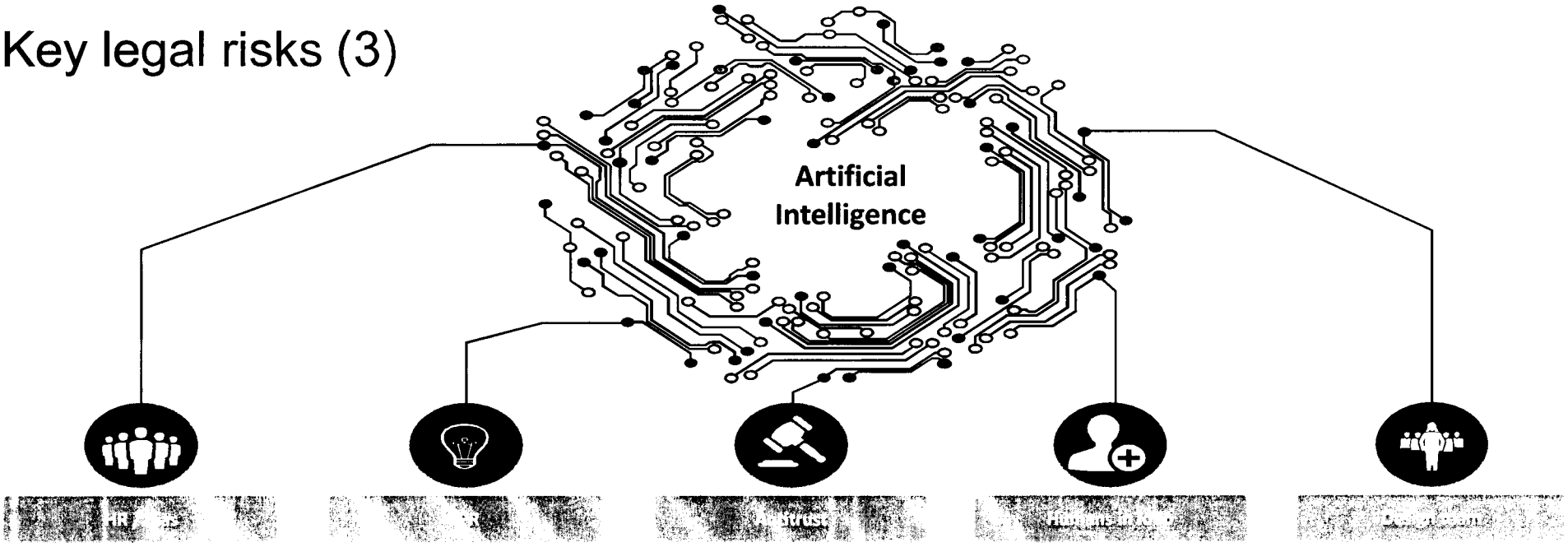
Are there any human rights or reputational considerations that the business ought to take into account?

Do any data privacy / data profiling issues arise in connection with the business's intended use of AI?

Is there a risk of cyber intrusion in relation to the AI, and who could incur loss as a result?

Has the business reviewed its insurance position in relation to its development or use of AI?

Key legal risks (3)



Does the intended use of AI raise any employment law issues, or issues of bias in relation to the business and the provision of its goods, services and facilities?

What steps does the business need to take to protect IP rights in connection with any AI it develops, and to protect itself from IP infringement claims in its use of AI?

Does the business's use of AI give rise to any competition / antitrust issues?

Is there any scope for the AI product to provide that a human has the power to intervene before decisions by the system are finalised in order to check them, or to override them afterwards?

Has the business made its AI product design/development team aware of the need to factor the issues above into the pre-product design phase of an AI project?

What is the impact of AI on IP?

- **IP protection for AI**
  - Layered approach like other computer technology products and services
    - Copyright, trade secrets, data protection, brands, designs and patents
  - Clear agreements in place for suppliers, customers, and collaborators
  - Open research and licensing arrangements
  - Extract commercial value

What is the impact of AI on IP?

- **AI as IP creator**
  - AI can create new works and inventions. Who owns the IP rights in relation to machine-generated original works?
  - use agreements with clear terms regarding ownership of IP rights; and
  - IP law may need to be revised to address machine-created works in some jurisdictions
- **AI as IP infringer**
  - a machine may act or operate autonomously in a manner that infringes third party IP rights
  - If existing laws do not extend liability to a machine, then a related stakeholder (such as the owner, developer, operator, or another AI Supply Chain Participant) may be responsible.
  - Limitation on Liability? Indemnification?

- Any questions?

Thank you! Merci!

Maya Medeiros

maya.medeiros@nortonrosefulbright.com

Karlin, Michael

From: Miller, James (IC) <james.miller@canada.ca>
Sent: December 20, 2017 12:01 PM
To: Karlin, Michael
Subject: RE: Artificial Intelligence

s.19(1)

Hi Michael,

This is very informative and thanks for getting back to me quickly. I look forward to reading your white paper on this.

I'm sure as we move forward on AI in the new year we will be in touch more.

Thanks again

James

From: Karlin, Michael [mailto:Michael.Karlin@tbs-sct.gc.ca]
Sent: December-20-17 10:54 AM
To: Miller, James (IC)
Subject: RE: Artificial Intelligence

Hi James,

Thanks for reaching out. This is a crowded space for sure so it's good that you're mapping out who does what.

Broadly, our job is to ensure that federal institutions have the right policy framework to allow for them to use AI systems in the design and delivery of their programs. This includes Treasury Board guidance, but also asset standardization, and a review of institutions' AI projects through the GC Enterprise Architecture Review Board.

The first product is our white paper, which you are welcome to review and comment on here. In the coming year, expect our role to ramp up considerably, including the design of Treasury Board guidance for automated decision support systems (AI or process automation) as well as standards around data governance and the use of open source software. Most importantly, it will be important to ensure that departments have clear guidance with regards to the interpretability and accountability of algorithms used for services. In that, we work closely with stakeholders in Canada and internationally on the ethical design of AI so that our rules for federal departments will be rooted in cutting-edge research and international norms.

We are also tinkering with the idea of working with the NRC and a couple of departments on the development of some experimental applications that can advance government business in some way. Our Minister supports the idea, so we'll see where that goes.

Finally, we have a data science practice starting here at TBS, though in a separate branch. It's a touch early to say what their focus will be (I suspect the first year will be data structuring!) but I can comfortably say that some predictive analytics will be involved.

I hope this information helps. If you need anything else, let me know but if you need further clarification, I'd be happy to provide it tomorrow.

Cheers,
M.

Michael Karlin

Advisor, Artificial intelligence – policy and implementation
Chief Information Officer Branch
Treasury Board of Canada Secretariat / Government of Canada
Michael.Karlin@tbs-sct.gc.ca / @supergovernance

Conseiller, Intelligence artificielle – politiques et réalisation
Direction du dirigeant principal de l'information
Secrétariat du Conseil du Trésor du Canada / Gouvernement du Canada
Michael.Karlin@tbs-sct.gc.ca / @supergovernance

From: Miller, James (IC) [mailto:james.miller@canada.ca]
Sent: December 20, 2017 10:25 AM
To: Karlin, Michael
Subject: Artificial Intelligence
Importance: High

Hi Michael,

I'm James Miller and I work within the Manufacturing and Life Science Branch here at ISED Canada. I've been told you are the AI expert at TBS. We are doing a stock taking exercise on what various federal departments and agencies are doing around AI, whether its targeted programs, or initiatives or policies. I wanted to touch base with you to get some information on what TBS is doing relating to AI. Would you be able to provide me with some high level information on what TBS is doing relating to AI?

Our management want to see something before the holidays so if you could get back to me as soon as you can it would be appreciated. Again, we are not looking for deep details just high level information that you may have handy or off-the-shelf.

Thanks

James

James Miller

Senior Policy Advisor, Manufacturing and Life Sciences Branch
Innovation, Science and Economic Development Canada / Government of Canada
james.miller@canada.ca / Tel: 613-219-2860 / TTY: 1-866-694-8389

Conseiller principale sur les politiques, Direction générale de la fabrication et des sciences de la vie
Innovation, Sciences et Développement économique Canada / Gouvernement du Canada
james.miller@canada.ca / Tél: 613-219-2860 / ATS: 1-866-694-8389

Pages 26 to / à 29

are withheld pursuant to sections

sont retenues en vertu des articles

13(1)(b), 15(1) - I.A.*

of the Access to Information Act

de la Loi sur l'accès à l'information

TBS-SCT

Responsible Artificial Intelligence in the Government of Canada

Digital Disruption White Paper Series

Version 1.3
2017-12-28

1. Table of Contents

2. Version History

#	Date	History
Sections 0.1 – 0.10 –Early Concept		
0.1	July 21	Release
0.11	August 7	Release
0.12	August 17	Release
0.2	September 22	Release
0.3	October 2	Release
0.4	October 16	Release
Sections 1.0 – 1.X – Working Drafts for Broad Consultation		
1.0	October 27	First draft for open comment
1.1	November 6	New introduction, new section on security and user access control
1.2	November 21	New sections on inclusion and cybersecurity, revisions to sections on AI for policy, revisions to "evolving how government works," new box on anthropomorphism, risk test appendix removed - it belongs more in directive format than in this white paper.

To-Do List

Item	Completed
Expand discussion on intelligibility	
Infographic: closed-rule algorithm, deep learning, reinforcement learning applied to one example policy, showing potential guard rails and pitfalls (let's see if this works)	
Expand Appendix A - how would this functionally work?	Not

	completed - Appendix A was too detailed for this paper, and will be included in a policy instrument
Add section on data linkages and the ability to gain new insights - as well as privacy pitfalls (per Scott's comment)	
Add sub-section on inclusiveness (per Hallie's comment)	
Add section on third-party service concierges	

3. Executive Summary

Artificial intelligence (AI) is a term used to describe a suite of related technologies intended to simulate and enhance human cognitive capabilities, such as pattern recognition, judgement, vision, or hearing. Having first been conceived in the 1940's, AI has advanced rapidly in recent years due to a combination of vast quantities of data, new mathematical techniques, and inexpensive computing power. AI systems now underpin many of the consumer products that Canadian use on a daily basis, from curating what media content we consume based on our interests, to helping us navigate our towns and cities. There are real-world examples of AI systems operating vehicles, writing newspaper articles, or generating art, that challenge previous assumptions of the types of tasks that can be delegated to machines.

Just as AI systems are rapidly transforming the world around us, so too is it expected that AI will transform the way that government operates. Imagine virtual service agents assisting Canadians and businesses with completing routine transactions 24 hours a day, seven days a week. AI systems can monitor the status of industries to detect early warning of regulatory non-compliance. They can sift through, structure, and recombine vast stores of data to help government institutions understand the information that they currently have, in order to more intelligently design public policy. These technologies have the potential to guide the public service towards a future of greater effectiveness and responsiveness to the needs of society than was ever possible before.

While the power that AI systems may bring to government could be significant, they must be deployed in a responsible and ethical manner. AI systems often require "training" using datasets that are reflective of the problem needing to be solved. If these data were collected or tabulated in a way that carries bias, then the outcome will be AI recommendations or decisions that are

biased as well. Further, some AI systems currently operate as "black boxes," meaning that the decisions they make are difficult to audit or fully comprehend. In light of these limitations, it is important to understand where it is appropriate to deploy different types of AI systems, balancing the potential for gains in efficiency and effectiveness of government with the risk of misuse. Finally, although AI will afford institutions with new capabilities, institutions will need to apply a strong ethical lens to whether the technology should be deployed at all in certain circumstances.

AI is a capability that rests atop an expert and disciplined data science practice within institutions, as well as leveraging Canada's leading AI talent base. These systems will challenge how government institutions work, demanding a prioritization of good data governance practices, and requiring new skillsets of knowledge workers.

This paper proposes a set of seven principles that will be expressed in all future Treasury Board policy on the use of AI systems in government:

1. People should always be governed – and perceive to be governed – by people;
2. AI systems deployed on behalf of government should be trained to reflect the Values and Ethics of the Public Sector as well as Canadian and international human rights obligations; they should be used to reinforce these values where possible;
3. Organizations are accountable for the actions of AI systems, and should build systems that are auditable;
4. Understanding the need to protect privacy and national security, AI systems should be deployed in the most transparent manner possible;
5. Organizations should ensure that reliable contingencies are in place for when AI systems fail, or to provide services to those unable to access these systems;
6. AI systems should be developed in a diverse team that includes individuals capable of assessing the ethical and socioeconomic implications of the system;
7. AI systems should be deployed in a manner that minimizes negative impact to employees where possible, and should, where feasible, be created alongside the employees that will work with them.

4. Introduction

First it was chess, then Go, then poker. One by one, we have taught machines to exceed us in some of our most treasured – and complicated – games. These accomplishments showcased advancements in techniques achieved much faster than predicted, and were at least partially responsible for kicking off an era of massive investments and excitement in artificial intelligence. We have trained machines to mimic the outcomes of human learning and decision processes, such as adaptation, bargaining, and bluffing. With successive and public displays of computing prowess by the likes of IBM, Deepmind, or Facebook, and the rapid growth of a startup ecosystem, advances in AI have begun to dominate the press and capture the public's imagination.

While AI was originally conceived in the 1940's, over the past decade, these applications have been deployed in such variable and extensive ways that it increasingly drives the modern economy. AI has replaced humans on stock market floors[1] and in the management of multi-billion dollar hedge funds.[2] It assists with medical diagnoses and operates complex machinery autonomously. It has been applied to corporate process and workflow automation to increase efficiency of their operations. AI agents are beginning to use natural language effectively enough to interact with humans via intelligent chatbots. There is a very high likelihood that by 2025, AI will touch every aspect of modern society in ways both visible and invisible to Canadians.[3]

Since the 1970s, early investments in Canadian researchers allowed an AI industry to bloom here. The advances of Canadian pioneers in machine learning positioned this country as a global leader in AI research, development, and application. Budget 2017 committed $125 million to launch a Pan-Canadian Artificial Intelligence Strategy to support these clusters and attract the talent they need to maintain their advantage. Establishment of superclusters in Montreal, Toronto, and Edmonton has seen both the rise of world-leading research institutes as well as an ecosystem of AI startups that are internationally competitive and driving innovation.

---

[1] See example: http://www.bbc.com/news/business-34264380
[2] See example: https://www.theguardian.com/technology/2016/dec/22/bridgewater-associates-ai-artificial-intelligence-management
[3] A qualitative survey by the Pew Research Center of over 2,500 academics, policy analysts and corporate executives found broad consensus to support this prediction. While the study was American, respondents were international. See: Pew Research Center, "AI, Robotics and the Future of Jobs." Link: http://www.pewinternet.org/files/2014/08/Future-of-AI-Robotics-and-Jobs.pdf

Now, the Government of Canada is looking into how it can harness the opportunities provided by AI to offer novel and more timely services to citizens and other users,[4] as well as improve the effectiveness and efficiency of its operations. Federal institutions are working towards offering better user experiences to make their services easier to use, but these gains will not accomplish a frictionless service environment if the person faces weeks-long backlogs in having a benefit application processed. Especially in circumstances where work is routine, AI systems can work faster and often more consistently than humans performing the equivalent tasks, and will work over evenings, weekends, and statutory holidays. Their capacities for decision-making are not adversely affected by physical fatigue or the natural emotional and relational situations people face based on their natural makeup. AI systems can be deployed by service institutions to answer questions posed by users – as well as make eligibility determinations – in order to dramatically improve the response time of service.

On the other hand, when administrative tasks are complex and value-laden, it can be difficult to ensure that the actions of the AI systems align with the spirit and intentions of the policy being implemented. Working with complex social and economic systems is considerably more complex than a game of Go. How do we know whether an AI system is appropriately trained for its task, and that data is interpreted in a manner that is accurate and responsible? How do we know whether AI is making biased or prejudicial decisions? How can AI systems be coded to meet similar legal obligations as human public servants, such as the Charter of Rights and Freedoms or the *Privacy Act*, and who is responsible when they fail to meet these obligations? How do we teach it social, cultural, or geographical context such that it can make decisions in a nuanced fashion? How do we know the rationale behind the decisions of an AI system? What types of decisions should always require some form of human intervention? How do we know that the data on which an AI system is trained, which is sampled from real data about real Canadians, is kept secure and private once the AI system is in deployment? What are the workforce requirements in a post-AI world?

Governments worldwide are now grappling with the consequences of a technological development that is transforming service delivery across sectors. The United States, United Kingdom, France, the United Arab Emirates, China and Japan are just some of the jurisdictions that have undertaken high-level examinations of AI systems within their respective governments and on their economies writ-large. The Government of Canada has the opportunity to build on the brain trust of private sector and academic leaders in this field to position itself as a world leader in AI for policy development and service delivery. It has the opportunity to signal to all sectors that AI can be harnessed in a manner that is ethical and supportive of positive outcomes for Canadians without sacrificing the benefits of the technology.

While AI is undergoing rapid advancement, it is important that the policy, ethical and legal implications of the use of this technology to deliver government services be addressed

[4] This paper uses the term "users" to represent the diverse groups that use Government of Canada services including, but not limited to, citizens, permanent and temporary residents, and businesses. It avoids the term "client" to reduce confusion with the legal term.

methodically and with an understanding of this complexity. The service delivery opportunities are significant, as are the pitfalls.

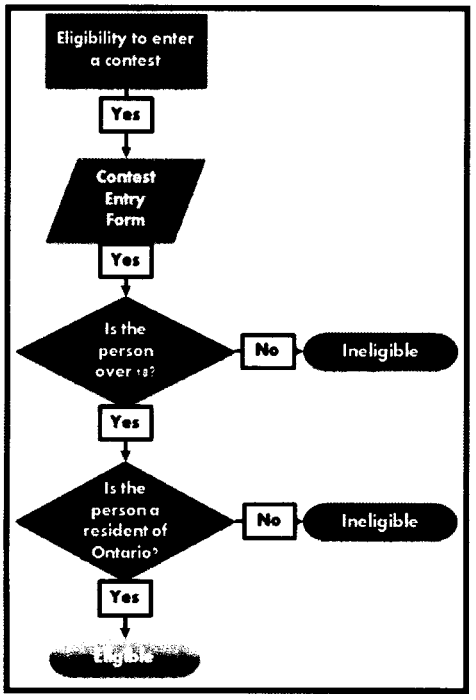
4.1. Objective of this paper and intended audience

The scope of this paper is limited to the specific use of AI applications by federal institutions for their own use only; it does not touch on the Government's response to automation in the private sector and its effect on society. This scope is broadly aligned with the mandate of the Treasury Board in its role in setting general administrative policy for federal institutions.

This white paper will examine the policy, ethical, technical, and legal considerations around the use of this technology within the Government of Canada. Its primary objective is to assist federal institutions by providing recommendations on how these systems should be implemented. The intended audience is therefore broad, from Deputy Heads or Chief Information Officers wishing to understand a significant new technology, to policy managers or service designers looking to apply AI to the programs or services that they provide. At the same time, it is intended to communicate to the AI development ecosystem in the academic and private sectors the use cases and policy considerations that are common in the federal government.

Throughout the paper, illustrative examples are used to show how this technology can be beneficial to users. Unless otherwise specified, these examples do not represent any existing plans of the Government of Canada and should be considered theoretical only.

4.2. Automation and Artificial Intelligence



Humans have always been intrepid designers of tools. From the scythe and wheel to the internal combustion engine and the computer, we have always designed tools to produce more from less. For most of human history this has led to technologies that have extended our physical capacities, but with the outbreak of the Second World War, humanity started designing tools that started to extend our cognitive and analytical capacities as well, such as memory, attention, judgement and decision-making. In a sense, we started designing brains for our tools.

We eventually designed tools that took over tasks for us completely. Automation has been a hallmark of industrialization since the robot Unimate was deployed in

a New Jersey GM plant in 1961 for hazardous die casting, not just for physical tasks, but for analytical ones as well.

Behind the automated processes that drive the 21st century economy are a series of logical instructions known as algorithms. Like a recipe, algorithms are processes that inform a machine how to perform a specific task. They can often be broken down into a series of decisions that are defined by the programmer; such as "is the individual over 18 years old?" or "is the individual a legal resident of Ontario?" The output is decided based on these decisions. The rules of these algorithms do not change unless programmers decide to change them. Closed-rule algorithms are used in the support of decisions widely in the private and public sectors today; for example, the Canada Revenue Agency uses closed-rule algorithms to support tax processing, with the rules defined by legislation and regulation.

Enter Artificial Intelligence

While it was the eminent British computer scientist Alan Turing that first conceived of "the thinking machine," the term "artificial intelligence" was coined later in 1956 by the American computer scientist John McCarthy to describe "the science and engineering of making intelligent machines." As technology has evolved, AI has grown to become a term that includes a broad spectrum of related technologies that seek to imitate and enhance aspects of human intelligence, such as vision, identifying patterns in information, or understanding language. In a sense, AI is when computers do what only humans could before. The term is used to describe applications as innocuous as a system that recommends books to read, to fictional advanced human-like intelligence capable of everything a human is. As such, there is no single, internationally-recognized definition for AI, and the term may mean different things to different people.

The development of **machine learning** was a critical milestone. Machine learning is a method by which algorithms can be trained how to recognize patterns within information, and the ways in which data interrelate. For example, a learning algorithm that recommends books based on your purchasing history provides better recommendations as you purchase more books. It does this without a human on the backend needing to adjust the programming instructions. If that algorithm had access to your browsing history as input data - and assuming that it was programmed to know what to do with that data - its recommendations may improve even more because it begins to "know" your tastes better.

Machine learning is by no means the only application of artificial intelligence. **Natural language processing** allows computers to parse meaning and context out of written text. This is used extensively, for example, in legal analysis software to derive insights from large volumes of text. **Machine vision and hearing** provide machines with the capability of structuring, and using, typically unstructured data such as imagery or sound. This is used in a diverse range applications, from autonomous cars "seeing" obstacles to smartphone applications that can identify a song played in public.

Either one or a combination of these techniques underpin many of the private sector digital services that people use regularly worldwide. Major social networking platforms, media platforms, and smartphones all run machine learning algorithms that provide services such as navigating traffic or curating news. It is not necessary to use machine learning in all approaches to automation; for applications where rules are precisely defined (such as the example above), a closed-rule algorithm is sufficient for the task.

Early experiments have existed since the late-1950s to show how machines are capable of learning and self-improvement. Today, researchers and developers have access to powerful and inexpensive cloud computing resources, parallel computing, as well as profoundly more data. Smartphones and the sensors located within them, coupled with the popularity of social media and internet culture means that a typical person produces a bounty of harvestable data every day - even when they are sleeping.[5] As a result, the development - and implementation - of AI has progressed rapidly in the last ten years. As the Internet of Things connects common consumer products and appliances to the internet, the data points that we generate in our day-to-day lives will likely grow exponentially.

This ability to capture and use data in unprecedented ways has had a direct impact on the development of AI because of these technologies' need for sufficient quality and quantity of data. Think of AI as a very sophisticated engine; without data to fuel it, it can't propel the vehicle. Data needs to be available in sufficient quantity, they need to be relevant enough to the task at hand, they need to have been collected and described in a manner that is free of bias, and they need to be in a format that is readable by a machine. Despite addressing AI, much of this paper is devoted to issues surrounding data rather than the instructions precisely because insufficient quality and quantity of data can render the most expertly-programmed AI useless - or worse - harmful.

We are now at a point where machine learning can enable AI not only to replicate many human tasks - it can come close to surpassing our effectiveness at certain tasks, such as recognizing subjects of images,[6] or reading lips.[7]

Advances in techniques

There are many approaches that developers take to AI; for example, **deep learning,** a branch of machine learning, has been used extensively in modern private sector services. While many deep learning algorithms use labelled data, it also brought the capability of using unstructured data such as audio or visual data, allowing the system to extract features of information on its own.

---

[5] For example, by using an app that monitors sleep time and quality.
http://ns.umich.edu/new/multimedia/videos/23822-smartphones-uncover-how-the-world-sleeps
[6] Based on 2017 results of the University of Washington MegaFace challenge:
http://megaface.cs.washington.edu/results/facescrub.html
[7] Based on LipNet results. See:
https://www.technologyreview.com/s/602949/ai-has-beaten-humans-at-lip-reading/

There have been significant advances in **artificial neural networks** in recent years. Inspired by the human brain, neural networks are composed of artificial neurons, which receive data individually and calculate outputs independently, allowing a complex problem to be broken down into millions of simple problems and then reassembled as one answer. As the network is provided more data, it can identify new and complex relationships in data, much like how the human brain forms synapses. This complex relationship is encoded in the weights, learned during model training, that connect the neurons in the neural network.

For example, rather than just learning what a bear is based on analyzing millions of images tagged as "bear," a deep learning AI can extract features from images of a bear on its own. Humans do that as well; we learn a bear's size and shape, where a bear may be found, typical colours of its fur and its family structure. That way, when we see an image of a bear that we have never seen before, we can infer that what we are seeing is a bear based on understanding its components.

The complication of deep learning is that it is not always possible to have access to massive data and to understand the importance associated with different variables of the problem. Using the above example, it is very difficult to understand whether an AI neural network considers fins as important than scales in determining whether something is a bear or not, both because the network is complex, but also because as the network is exposed to more examples of bears, this weighting may change. This process is often reliant on very large volumes of data that are broadly representative of the world within which the system will operate; for example, an autonomous vehicle trained exclusively in the UK could not be deployed in Canada, where driving is on the opposite side and some rules differ.

Another approach is, reinforcement learning; this is a subset of machine learning whereby machines are trained by being rewarded for desired outcomes and punished for undesired ones, similar to how we train bears to play fetch.[8] Rules are provided to the algorithm as to what it must do to earn a reward; for example, if the bear brings the ball back, it will get a fish to eat. The bear will not receive a fish if it does not bring the ball back. Reinforcement learning is especially useful in situations with well-defined outcomes, for example games and puzzles.

Reinforcement learning algorithms can be trained in advance using simulations, but they adapt more quickly once able to interact with its intended operating environment. However, they need clear definitions of "right and wrong" - outcomes that are desirable or undesirable, and the choices of those definitions are laden with values.

The choice of methodology will matter depending on the problem needing to be solved.

[8] Please do not attempt to play fetch with bears.

4.3. Narrow and General Intelligence

Whereas you are a multifaceted individual with a number of potentially unrelated interests, AI is often targeted for a single objective or task. This is known as "narrow" intelligence; while it can excel at one task – even surpassing a human – it cannot learn a second task without being explicitly targeted to do so. For example, while you may be a software engineer that speaks four languages fluently and is an amateur chef, an AI system trained to identify high-risk travellers cannot simply choose to learn to translate languages. This is because AI is software and does not have agency.

While research is underway to determine whether AI can achieve general intelligence, this achievement is still highly theoretical. A generally intelligent AI brings with it significant policy implications as well, but this paper will focus on the implications of narrow AI.

AI is software, not an organism

For decades, science fiction has introduced AI characters – whether in robot or incorporeal form – to the social consciousness. The popularity of characters like HAL 9000 or C-3P0 may cause us to ascribe some degree of personification to AI. While it is designed to mimic human intelligence, the "learning" and "understanding" that a machine undergoes is different than the biological processes that we humans rely upon.

This paper refers to AI using humanlike semantics from time to time because it is a helpful way to communicate technical concepts, but it is important to remember that fundamentally, AI is software, not a conscious being, and should not be ascribed agency over its actions. Doing so removes the accountability of an organization over its software.

5. AI for Smarter Government

AI is not a technology looking for a problem; it is a suite of tools with the potential to help the GC deliver services more effectively, design policy more responsively, and potentially enable an entire suite of new capabilities in designing policy and delivering services. As the set of applications is diverse, its potential impact on the public sector is wide-ranging. Institutions have been examining applications that can be organized into three interdependent themes:

1. Applying AI to the delivery of services to the public
2. Applying AI to help design policy and respond to risk
3. Applying AI to the internal services of government

5.1. AI for the Delivery of Services to the Public

End-to-end digital self-service is the norm throughout much of the private sector service spectrum. The ability to access the entire continuum of the service from application to delivery without the need for a paper form, or for the user to have to interact with a service agent, is typical. Ideally, the service experience from authentication to application to receipt of benefit or issuance of payment should be a seamless process that does not require the use of a phone or visit to a service centre unless chosen as the preferred way to receive service.

The government has decided to prioritize the development of digital services. Phone and in-person channels are inherently less convenient for users, as opening times are restricted, require waiting on hold or in line, or involve travel times. For individuals and businesses alike, lengthy wait times, or the requirement to access services during business hours can lead to an unacceptable loss of leisure time or productivity. Assuming that the digital service offering is understandable, convenient, and accessible enough for someone to want to use it, there is incentive for all parties of a service transaction to want to move to the digital channel. According to the Canadian Radio-Telecommunications Commission, broadband access in Canada will likely reach 90% as soon as 2021, and digital services will be more within reach for the vast majority of Canadians.[9]

Even if all services were provided digitally, as of today, there may be services by which some will elect to use other channels. There are some complex or sensitive needs that may demand more nuanced or personal service provision. Some people simply may feel more comfortable raising their issues in front of another person. In these circumstances, people may use an alternative channel such as phone or in-person if these are accessible to the individual. Even in these cases, AI can empower services by providing faster decisions, or tools that provide an overview of the individual's sentiment during the progress of the call.

More intelligent digital tools interacting directly with a user can play a role in keeping them on the digital channel. Smarter search and chatbots are capable of parsing natural language into searchable terms, accessing information located in FAQs, manuals or even specifically-identified internal documents and reply to the question in a way the user can understand. With additional information and user feedback, these tools will continuously improve at this task without the need for direct human intervention.

5.1.1. Smarter Search

Building a website targeted at millions of people presents a challenge; people interpret information differently, and may have different expectations as to where information can be found. Usability testing can help understand how people are interpreting information on a

[9] CRTC http://www.crtc.gc.ca/eng/internet/internet.htm

website, but advances in natural language processing (NLP) make the task of finding relevant information much easier than it used to be.

NLP technology parses natural language into underlying meaning, which then can be used in service of some task. For example, if a user loses their job, rather than having to look up Employment Insurance specifically, search for "I've lost my job" and see results that are relevant to that request. Over time, the application learns the relevance between search statements and the services that people are looking for. This is superior to older search methodologies, which would literally scan for the statement "I've lost my job" in web content. Over time, the algorithm will learn more patterns and do a better job at understanding what users want. NLP search functionality is widely used in the private sector today.

5.1.2. Chatbots

Chatbots are virtual user service representatives that offer capabilities of searching for information, or escorting a user to the right webpage. They work similarly to NLP search, but add a layer of interactivity and personalization.

The capabilities of an AI chatbot can be scaled up over time to provide expansive levels of user care as it gains experience and improves the way it manages information. It can offer responsive services, answering queries related to services passively. Eventually it can expand to become more navigational, offering hints, advice, or step-by-step instructions more reflexive of where a person is in the continuum of their service experience. Eventually an AI can be capable of actually executing instructions, such as accessing and pre-filling a form based on natural language.

A chatbot may be offered to clients embedded in your webpage, or within another platform where your users are commonly found, such as SMS text messaging, Facebook Messenger, WhatsApp, Twitter, or Slack. This technology has advanced significantly over the past five years and is expected to continue its rapid advancement for the next decade, for providing both external and internal services.

Chatbots offer a diverse opportunity to provide services to users. Chatbots help filter routine questions away from human service agents so that they may focus on helping users through complex or distressing cases, or cases where a user is uncomfortable relaying their circumstances to a machine. They may also assist with public consultations on policies or programs, by being able to ask follow-up questions and react to user feedback in a much more nimble fashion than a survey.

This technology has been deployed successfully in the public and private sector. The United States Citizenship and Immigration Services uses a chatbot named Emma to answer users' questions and provide a pre-check for eligibility. Emma not only answers questions, but provides navigational services; the search query "I've been offered a job in the US" not only provokes a

response from Emma, but brings the user to the "Working in the United States" site. The bot is trained in English and Spanish. Another bot, Sgt. Star, is deployed by the US Army to answer questions to prospective recruits.

Institutions looking to deploy chatbots will need to ensure that there is training data available for the bot to learn the appropriate terminology for the service. This data can include previous interactions with clients looking for the service in question, whether emails, chat logs, transcripts from phone conversations, or social media. Ideally the datasets would include data on the outcome of the service interaction as well to ensure that responses to questions are those that actually satisfy clients.

Chatbots have limitations. As described above, conversations carry a lot of information outside of the basic text. Emotional queues or the use of sarcasm and humour can quickly confuse an AI conversational agent, or teach it bad behaviour. While they are adept at managing basic questions, a lengthy, interactive conversation is not possible at this time. Some chatbots provide a user with a defined set of potential inputs to reduce errors in the conversation, which results in a more scripted interaction. This can be useful for quickly helping users find the information they need, although scripted interactions quickly become difficult to control as the scope of the bot's responsibilities increase.

An additional benefit of chatbots is their ability to structure data through a standardized approach to collection. Through interactions with users, a chatbot can help reduce spelling errors, inappropriate entry of dates and addresses, etc. This improves overall data quality, which in turn could help eligibility determination.

Chatbots offer transactional capability as well, merging the functions of both a virtual front-line service agent and the application form by collecting information directly from the user or their file in the institution's Customer Relation Management software.

It's important to remember that a user interacting with a chatbot may ask questions that are well outside the scope of its expertise. Users may disclose important personal information even when advised not to; they may even require immediate emergency assistance. In such a circumstance a human would be guided by a mix of their training and their own moral compass, but machine intelligences would need pre-programmed responses for these situations.

Just like a human agent, a chatbot needs to be treated as an agent of the organization, which means that the information that it provides must always be accurate and up-to-date. Learning chatbots may provide advice to Canadians and, like humans, sometimes make mistakes. For example, a chatbot may give a person the wrong form or provide them an incorrect deadline. Chatbots that are designed to actually replace a form through conversational means may misinterpret input and submit incorrect information.

There have been significant and swift advances in chatbot technology, but despite these advances, it is a long way from flawless. In the future, bots have the potential for replacing forms as a way to collect information from users. They may even emerge to become the primary service delivery platform. Assuming that they have access to the widest range of information possible, bots can theoretically inform a user about any service in any institution with an almost expert like knowledge, far surpassing the ability of one individual's recall.

Finally, there are those in Canada who do not have access to reliable broadband internet, and may not in the near future. It is important that institutions continue to cater to these users and do not solely rely on chatbots for front-line services.

Is your institution ready for a chatbot?

When determining whether to deploy a chatbot, an institution should be able to answer the following questions:

Is there a clear business driver for the chatbot? Does your institution receive a high volume of routine inquiries?

Are the most common inquiries known and are data available to answer them?

What can be automated without taking away from the user experience and satisfaction?

What is the sensitivity of the information that the chatbot will likely receive or relay?

Will the interaction be an entirely scripted one, or allow the user to ask open questions?

Will there be an escalation process to a human live chat?

Can interactions be stored in your CRM? Will it enable engagement across other channels (e.g. email, phone, in-person)?

5.1.2.1. User Experience Considerations

The GC has a wide policy and service landscape; if chatbots speaking to these policies and services offer interaction experiences that differ significantly, then users' acceptance of this technology can suffer and benefits will be unrealized.

A chatbot should not be used as a substitute for good discoverability of information on a website; it can add supplementary information or clarification to a user, but should not be seen as to replace the need for a well-designed site.

Institutions looking to deploy chatbots should consider the following:

- Chatbot conversations should be introduced with a brief privacy notice that is compliant with the *Treasury Board Standard on Privacy and Web Analytics*. This notice should provide a link to a page with more information on the information collected in the course of the conversation, including any metadata, for example: time and date, duration, whether the conversation was ended by the user or the agent, whether and when the discussion was escalated to a human, etc.
- Whether the bot is able to provide a professional tone as a representative of the Government of Canada. Machine learning chatbots may learn language that is potentially unprofessional, abusive, or harassing if exposed to sufficient examples. Where possible, institutions should work with vendors to prevent them from learning this behaviour, whether using a keyword blacklist, or other methodology. It is important to be continually monitoring chatbots' performance in this regard.
- Whether to use an avatar, which is a personification of the chatbot. Visual avatars that express some emotional range improve users' belief in the competence of the virtual agent.[10]
- The question of whether or not a chatbot should be gendered as male or female - or, for that matter, anthropomorphized (meaning: made to appear human - deserves close attention. It is unclear whether the use of a female gendered "assistant," could serve to perpetuate false, misleading and ultimately harmful cultural stereotypes about the status of women. To avoid a misstep in this sensitive area, some organizations have made the proactive decision to characterize their assistants as androgynous, such as Capital One's *Eno* and Sage's *Pegg* or non-human, such as Google's *Voice Assistant.*[11]
- Institutions should be mindful that people in rural or remote locations may encounter latency that will affect their ability to respond to the chatbot's queries. It's important to ensure that response times from the user are permissive.
- Chatbots must be accessible and meet accessibility standards and requirements of the GC. They should use plain language so as to be understood by users with varying

[10] [6] Demeure, Niewiadomski and Pelachaud, "How Is Believability of a Virtual Agent Related to Warmth, Competence, Personification, and Embodiment?" *Presence*, October 2011. Link: http://www.mitpressjournals.org/doi/pdf/10.1162/PRES_a_00065

[11] For more on Eno and Pegg see: https://www.accountingtoday.com/opinion/the-tech-take-the-genderless-face-of-accounting-bots
For more on Google's Assisant, see: https://www.engadget.com/2016/10/07/google-assistant-desexualize-ai/

levels of education or comfort with Canada's official languages. It is also important that chatbots be able to be read by screen readers, or are able themselves to communicate vocally, for persons with visual disabilities.

- Users should be provided with a clear escape from the conversation. If a user finds that a chatbot is no longer useful, or is incapable of answering their query, there should be a clear means to transfer the conversation to a human agent (if available), or to send email correspondence. Additionally, if a chatbot has answered a query and the user has ended the session or refrained from answering another question, the chatbot should politely end the conversation.

Towards a national chatbot training corps?

The more that a chatbot experiences, the more likely it is to develop a rich knowledge of its area of focus. Exposing the bot to as many knowledgeable trainers as possible will improve the bot's ability to learn variation of language. This comes with risk; the more trainers that are added to a training process, the higher the possibility of exposing the bot to bad language or inaccurate knowledge.

One way to balance these considerations may be to harness the collective intelligence of Canada's various community groups devoted to developing technology for the civic good. In addition to government trainers, organizing a voluntary "civic bot training corps" may be a way to help chatbots amass the right knowledge and understanding of language while minimizing the chance that it will be hijacked by people with malicious intent.

5.1.3. Automated Decision Support

Improving users' experiences when interacting with government services is important, but the benefits of this work are lost if the wait time to receive eligibility decisions on services is too long. Part of service excellence is cutting wait times, and AI can play a role.

To start, AI can be applied to electronic forms – both user-facing and back-end – to help ensure that data entered meets your institution's standard of quality. This modest application can greatly assist your institution's ability to use the data for decision-making later on.

Processing service applications requires that an analyst review application information, verify to see if it is true and believable, and checking if the information that has been submitted meets the program's eligibility criteria. This process can take time, both due to the amount of information collected as well as the limitations on resources.

By using appropriate program-related input data and a model to test inputs against rules, such as legislative or regulatory requirements, an automated system may be able to process eligibility decisions faster than and as well as a human in many circumstances. This allows eligibility analysis to be processed outside of core work hours, for data analytics to be gleaned and acted

upon promptly and organically, and for patterns to be established so that particularly complex or unexpected applications can be investigated more thoroughly. Strictly speaking, this approach can be done without the use of AI, as the rules themselves are strictly defined by the institution.

This level of decision automation has been tested and deployed in private sector settings for over a decade. Insurance and financial sectors have been pioneers in decision automation to improve service response times and to increase fraud detection. These sectors have similar challenges to governments: mission-critical systems with many dependencies, limited budgets and competing priorities for IT development, and a desire to maximize transaction throughput and minimize fraud.[12]

What if the system was designed in such a way that humans did not choose the eligibility criteria at all, but allowed a machine to determine what applicants should be eligible based on desired outcomes? For example, imagine a hypothetical program that provides small grants to exporters. Rather than have the program experts select the eligibility requirements themselves, an AI system analyzes similar firms in similar industries, and determines the likelihood of success following the grant. Of course, choosing the metrics that define "success" remains the responsibility of the program, but the criteria may vary. Perhaps there are different predictors of success for different sectors, or predictors that human analysts missed.

This approach has the potential to provide services with more effective outcomes, but brings challenges. For example, criteria are often enshrined in legal authorities. If there is a challenge to the decision, the institution would require to show what criteria were used to make the decision, something that might be difficult to show using current technology. This issue is further elaborated below.

Many government services have existed for decades; assuming there is high-quality, machine-readable data available, there is a significant volume of potential training sets to train AI how to process eligibility. By showing AI examples of successful versus unsuccessful applications, it can determine the necessary patterns to extend this reasoning to a new application on its own, effectively mimicking the experience of a human. For this to work, institutions need to have data on the outcomes of services in a format that is readable by machine.

5.1.3.1. Appropriateness of Automation

Should a service be automated completely from end-to-end, or should human intervention and approval always be required? The suitability of an automated system to deliver end-to-end services must be analyzed on a case by case basis. Much depends on the type of decision being made and the amount of discretion that any particular decision requires. Departments will have to carefully consider:

[12] See McKinsey report, "Automating the bank's back office," Link: http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/automating-the-banks-back-office

- Whether they are acting within lawful boundaries;
- Whether additional authorities are required;
- The procedures and mechanisms to be implemented to ensure transparency and to be able to document how a decision was reached, especially when such a decision affects individual rights and privileges and involves the exercise of discretion.

A "human in the loop" may not be straightforward in overturning machine decisions. Unless they are specifically instructed to, human officers will need to bring themselves to question the authoritativeness of the machine recommendation. Enough information would have to be provided to the human - both from the original input data such as a benefit application - but also around the rationale behind the decision. The human analyst should be required to document why the machine recommendation was not followed. Machine decisions flagged for human approval or overturning would themselves have to be monitored to ensure that there is no internal conspiracy or mismanagement.

The Government of Canada provides a diverse set of programs and services across over 140 federal institutions. Some of these programs and services are critical to the fundamental well-being of people, the economy, and the state; others are less. Should the same rigorous governance and accountability measures be required for non-critical programs as critical ones? Can we classify programs and services into risk categories to better target governance to be proportional to risk?

As TBS prepares guidance on how institutions can responsibly introduce automated decision support to their organization, it will develop a tool by which institutions can assess the degree of automation that is appropriate for their program. Guidance on governance could then be linked to the risk score.

5.1.3.2. Transparency and Recourse

How much information should be provided to users on the decision-making process? The ability and need to explain algorithmic decision making requires a delicate balance. On one hand, transparency builds trust and social acceptance, and provides users with information with which they can challenge decisions and business processes. On the other hand, providing too much information to the public can open a door to malicious manipulation of the algorithm.

Users should be notified in advance of submitting an application that it will be processed by an algorithm, along with a link leading a webpage with accessible, non-technical information on the decision-making process. This information should include a description of the sources of data used to make the decision, and links to recent system performance audits.

Further research is required to determine whether users should be provided an opportunity to opt out of automated decision making in advance of applying for a service. On one hand, this provides users with more control over how their personal information is handled. On the other

hand, designing systems for this to occur may be impractical and expensive. Regardless, in the event of a negative decision, users should be provided with an opportunity to have their application revisited by an informed human case assessor.

Further research is also required on what information institutions should provide on the design and functionality of AI tools (algorithms, logic, decision making rules), understanding that algorithms may be manipulated with too much of this information.

Regardless of the methodology used, it's important that institutions only automate a process when they have obtained a high level of confidence in the decisions that it is making in a test environment.

5.2. AI to help design policy and respond to risk

What if we were more accurately able to predict migration flows, forest fires, or the impact of an aging population? What if we knew in advance which ports of entry would be more likely to encounter contraband, or which consumer products might be more susceptible to recall? Existing analytical models have already given the GC the ability to better understand certain social or environmental outcomes to policy, but with new methods able to identify patterns in data that perhaps humans were previously incapable of doing, we may be able to make more precise and informed predictions than ever before.

Governments work with big problems. We work in an environment often marked by complex, interdependent systems, where small policy changes can result in massive impacts among a population or the economy. If we can use data to predict the impact of our work with greater precision, or to understand future pressures on social or economic programs, then we can respond more efficiently and ensure that regulatory resources are focused on the highest risk elements of their industries.

Using both structured and unstructured data sources, institutions can enhance their ability to understand what is happening in society and the economy, both in Canada and beyond. This will allow for more effective regulation of industries, as well as more informed policy planning through the use of simulation. The ability to combine even anonymized data sets across institutions in real time may be able to provide policymakers with new insights as to what is causing certain outcomes in society.

There are some limitations to this approach. Predictions are extrapolations of patterns that appeared in the past; while access to vast data sets brings greater opportunity to predict in a complex system, AI can't make truly novel predictions, because the past is not necessarily an indicator of the future. Like all AI systems, the right quantity and quality of data will need to be accessible to make accurate predictions. There is also a risk that predictions are made using data that has been collected in a way that is biased or not fully representative of the world that we live in; this issue is further discussed below.

Already, many federal institutions use a method to describe and compare the degree of risk involved with providing a service to a user. This "risk scoring" technique can be an efficient method to associate an administrative action with risk. To date, this has most often been accomplished using methods that require institutions precisely defining what risk is in their universe. These "closed-rule" algorithms, while not AI, are a form of automation that has shown to be service-enabling by reducing compliance and enforcement burden on lower-risk users.

5.3. Applying AI to the internal services of government

A professional public service is supported by intuitive and efficient internal services. Some of these services directly service Canada's democratic institutions, such as access to information or responses to the questions of parliamentarians. Others are in place to ensure that the public service itself is functioning smoothly, fostering a positive work environment and securing public assets.

5.3.1. Information Management

From white papers such as this one, to briefing notes, presentations, data sets, and other analysis, the GC is sitting on a vast trove of data, structured and unstructured, tagged and untagged. Traditional means of using this data has been limited to specific, machine-readable formats, but advances in semantic analysis have unlocked the potential for information in text format to be mined for insights as well. Now machine-usable information can be gleaned from text, audio, or video.

This technology can be used for a variety of applications, such as analyzing social media reaction to government policy or events; summarizing past briefings or approaches to maintain institutional memory; or automatically creating documentation trails for internal audit purposes.

The power behind these applications offers the promise of AI eventually providing virtual librarian services. With properly structured and tagged text data, a policy analyst will be able to more easily sort through and summarize past approaches to a problem, or find what is being done in other institutions. Having a smarter content management system understand what an analyst is looking for will help ensure that policy options are driven by data and that corporate memory is retained, leading to greater institutional wisdom.

5.3.2. Automated Content Generation

Over the past several years, products have entered the market allowing for content, be it text, audio, or visual content, to be generated automatically. Systems have been deployed in the private sector to automatically produce newspaper articles, blog content, or marketing copy. One notable example of this technology has been at the Associated Press newswire, which is estimated to be able to generate 2,000 news articles a second. After several months of training,

configuration, and maintenance, the system is now able to post stories without any human intervention at all. The "AI journalist" is capable of doing this because a) there was a dataset large enough for the computer to extract best practices, and b) most of these reports contain only factual information, with limited nuance.

There are potential applications for the business of government. This technology can likely be adapted to a number of government documents that are produced on a regular basis in large quantities that are often factual and follow a certain formula or template. While certainly incapable of making normative considerations, this technology can be useful to summarize and compare. For example, it would be able to write Ministerial correspondence, background sections of briefing or meeting scenario notes, background of Question Period notes, etc. This would allow human public servants to focus on analysis, policy lenses, considerations, and strategies for next steps.

5.3.3. People Management

AI is transforming the discipline of human resources management, whether to gauge and optimize productivity, or to match individuals to suitable jobs. The ability to scan through the information of thousands of candidates using a more precise and insightful method than static keyword searches can potentially lead to more effective hiring decisions. Understanding the skills and credentials of effective and ineffective employees can provide insight as to the attributes of an ideal candidate. This can improve overall organizational effectiveness, but also help an individual find a job they may be ideal for but may lack traditional qualifications.

Another HR application of AI is performance assessment and management. These tools measure an employee's effectiveness against certain criteria, such as delivering on projects or replying to stakeholder inquiries. Using these tools, a manager is able to have a dashboard of the productivity of employees and the current status of their projects.

These tools can bring ethical risks and must be deployed with great care. For many of these systems to work properly, a continuous volume of data must be collected about a person's productivity. This is tantamount to ongoing surveillance of the employee, something that could cause harm to the employee's mental health.[13] Deep and persistent AI supervision of employees may contribute to the very anxiety that reduces their effectiveness at work, which in turn may hinder them from changing jobs. Furthermore, this system would have to reflect the changing context of a job, such as busy or quieter periods of work (i.e. in media relations), or jobs that produce work that is difficult to easily quantify (i.e. policy advice).

Additionally, identifying optimal productivity may fail in certain cultural contexts, as some employees may work differently. A veteran, indigenous person, or someone born abroad may choose to work different hours, or using different techniques, which while effective, may be

[13] http://onlinelibrary.wiley.com/doi/10.1111/ntwe.12039/abstract

difficult to measure. An AI trained only on employees of European descent may not effectively evaluate an employee that is not. The systems would be required to consider the diverse accommodations that may be required for employees with certain disabilities.

At the current state of technology, AI systems should be prohibited from making unsupervised decisions about HR. When AI is generating recommendations for management, it is very important that employees be made aware of them in advance if at all possible, and be provided with the opportunity to access the information collected about them.

5.3.4. Security and Access Management

AI can be applied to the way institutions provide, review or revoke IT system and building authorizations by establishing baseline normal behaviour of staff and learning when certain activities seem out of the ordinary. It can provide a better alignment of IT security with operations and reduce the number of ad-hoc requests for access to a system. This can reduce the workload of IT administrators, allowing them to focus on user needs that are exceptional.

AI-powered cybersecurity and access control can further assist by allowing the detection of user needs at a granular level within a very short time, allowing users to have permissions better suited for what their job actually requires. AI can also be used to optimize permissions in business continuity planning.

Finally, there have been advances in machine learning cybersecurity applications that are designed to identify threats earlier, including internal threats where a sudden change of behaviour raises concern. While AI offers great promise in cybersecurity, it should be viewed as a single layer of protection, and not a substitute for existing systems and processes.

6. Policy, Ethical, and Legal Considerations of AI

With all of the potential use cases offering to improve policy and services, enthusiasm for AI in government has been high. Unfortunately, improper application of this technology can lead to negative outcomes for users, from frustrating service experiences to being mistakenly denied eligibility for benefits.

While the use of AI offers a lot of promise in improving the efficiency of government, it is important to approach its use with a strong ethical foundation. Machine ethics have been debated for years, and the Government of Canada should learn from these groundbreaking discussions to ensure that this transformative technology best serves the interest of everyone living in Canada.

As these agents grow to operate in increasingly sophisticated spaces, they act on behalf of the Crown, and should be subject to similar values, ethics, and laws as public servants and adherence to international human rights obligations. Institutions should incorporate these ethical principles in their application of AI:

1. **People should always be governed – and perceive to be governed – by people;**
2. **AI systems deployed on behalf of government should be trained to reflect the *Values and Ethics of the Public Sector* as well as Canadian and international human rights obligations; they should be used to reinforce these values where possible;**
3. **Organizations are accountable for the actions of AI systems, and should build systems that are auditable;**
4. **Understanding the need to protect privacy and national security, AI systems should be deployed in the most transparent manner possible;**
5. **Organizations should ensure that reliable contingencies are in place for when AI systems fail, or to provide services to those unable to access these systems;**
6. **AI systems should be developed in a diverse team that includes individuals capable of assessing the ethical and socioeconomic implications of the system;**
7. **AI systems should be deployed in a manner that minimizes negative impact to employees where possible, and should, where feasible, be created alongside the employees that will work with them.**

The Government of Canada is committed to incorporating international norms and standards in ethical design when applying AI or any autonomous system. The first step to preventing negative outcomes is to understand what they are and how they occur.

There is no "average" Canadian; this country consists of a population diverse in background and circumstance. There will be users with unique challenges that will test the rigour and limitations of algorithms deployed by government. Institutions need to account for exceptions, minimizing cases that fall through the cracks, and providing recourse for the inevitable failures of the system.

6.1. Data, Bias, and Rights

Every field of data entered is an investment for the future. That data will be examined, validated, and manipulated individually and in aggregate possibly thousands of times in the cycle of their life. Traditionally, data entry was viewed as an input cost to be minimized by many federal

institutions, but as the world moves more towards data-driven decisions, organizations are centering data governance in their core operations. This has unfortunately revealed a lack of consistent quality in data holdings.

Many AI applications are only as effective as the quality and quantity of their input data. The first step for an institution wishing to deploy an AI application is to ensure that the necessary training data is available, representative of the problem that needs to be solved, is readable by machine, and that the organization has the legal authority to collect and use this data. It also means adopting a culture of good data practices, and investing in the people and systems necessary to create, store, protect, and use data effectively.

6.1.1. Prevention of Data Bias

AI systems are not neutral; they will learn the biases of its programmers and the datasets used to train it. While unintentional, this bias can have ramifications that could range from embarrassing to serious. Even data that is incorrectly entered or labeled can have knock-on effects that affect real people in real ways. This can particularly affect vulnerable populations, of whom data has been collected historically with varying quantity and quality.

The ability to distinguish, predict, and learn means that AI is able to think in a more abstract fashion than earlier forms of computing. To do this, AI needs to be trained with datasets and oriented towards preferable outcomes. Both the training process and the selection of preferable outcomes carry with it the bias of the humans that collected and tagged the data, as well as the programmers that designed the algorithm. The collection of some data can be imperfect due to social or cultural stigma; for example, suicides and sexual assaults in Canada are both underreported.[14][15] Even the choice of which datasets to use and which to reject may entrench bias into the decision, and can lead to different outcomes.

Without enough training, an AI will have difficulty achieving its task, or will do so in a way that could lead to misinterpretations of data. Data collected in a certain socioeconomic context will echo in the decision-making of algorithms. The responsible policy manager needs to ensure that this important context is added to the analysis, and that they understand potential ways that AI can interpret input data incorrectly. Even controlling for certain variables won't necessarily protect from bias, as it can be derived from other, correlated variables; for example, excluding ethnicity from analysis won't necessarily protect from bias if the system can infer ethnicity from another variable such as a name.

The results of data bias can be highly problematic. As AI applications are more widely dispersed throughout society, a number of these unintentional but notable biases have been uncovered. For example, an algorithm used to predict crime in the United States has been shown to reinforce discriminatory policing because the crime data upon which it was trained was

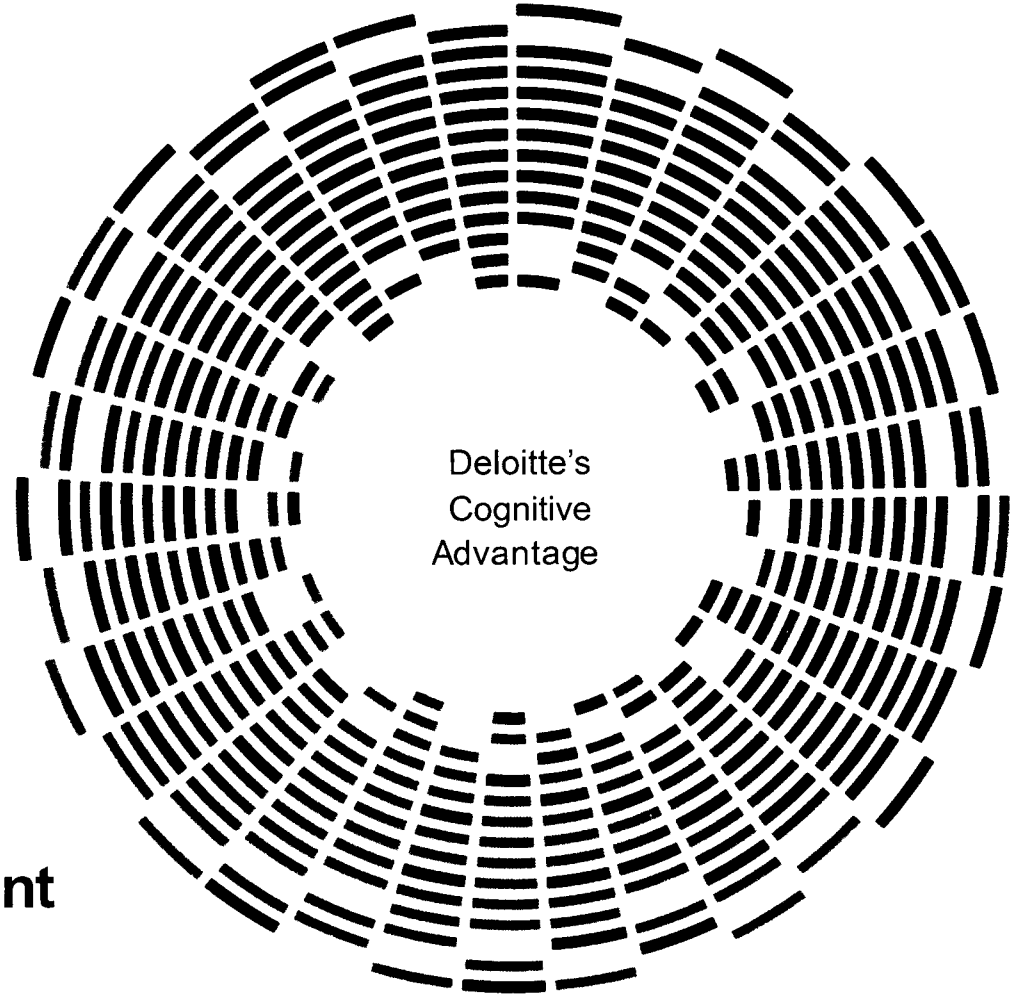
[14] https://www.statcan.gc.ca/pub/82-624-x/2012001/article/11696-eng.htm
[15] https://www.statcan.gc.ca/pub/85-002-x/2017001/article/14842-eng.htm

Pages 56 to / à 58

are not relevant

sont non pertinentes

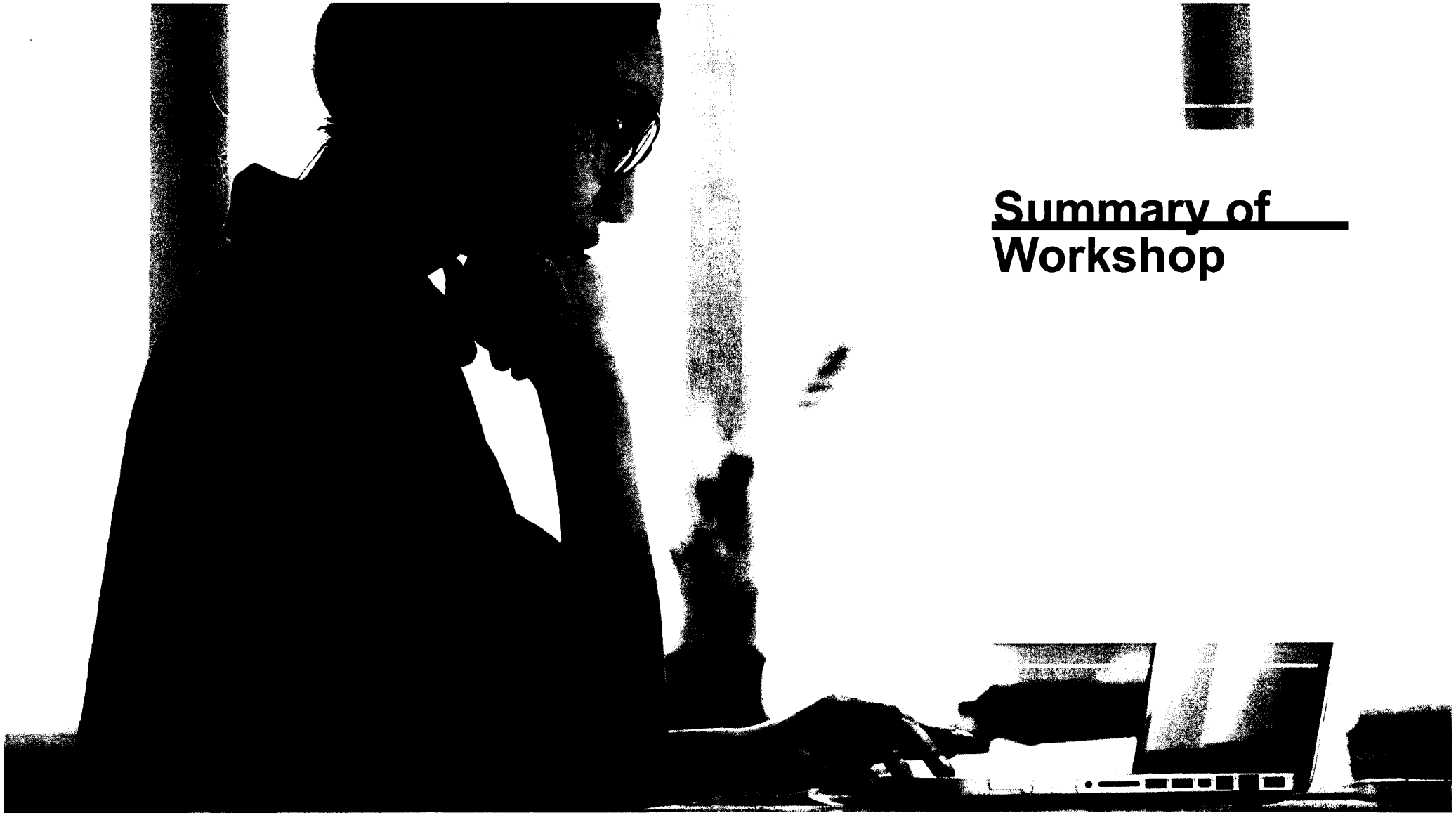
Deloitte.



FEBRUARY 2018

Output Document

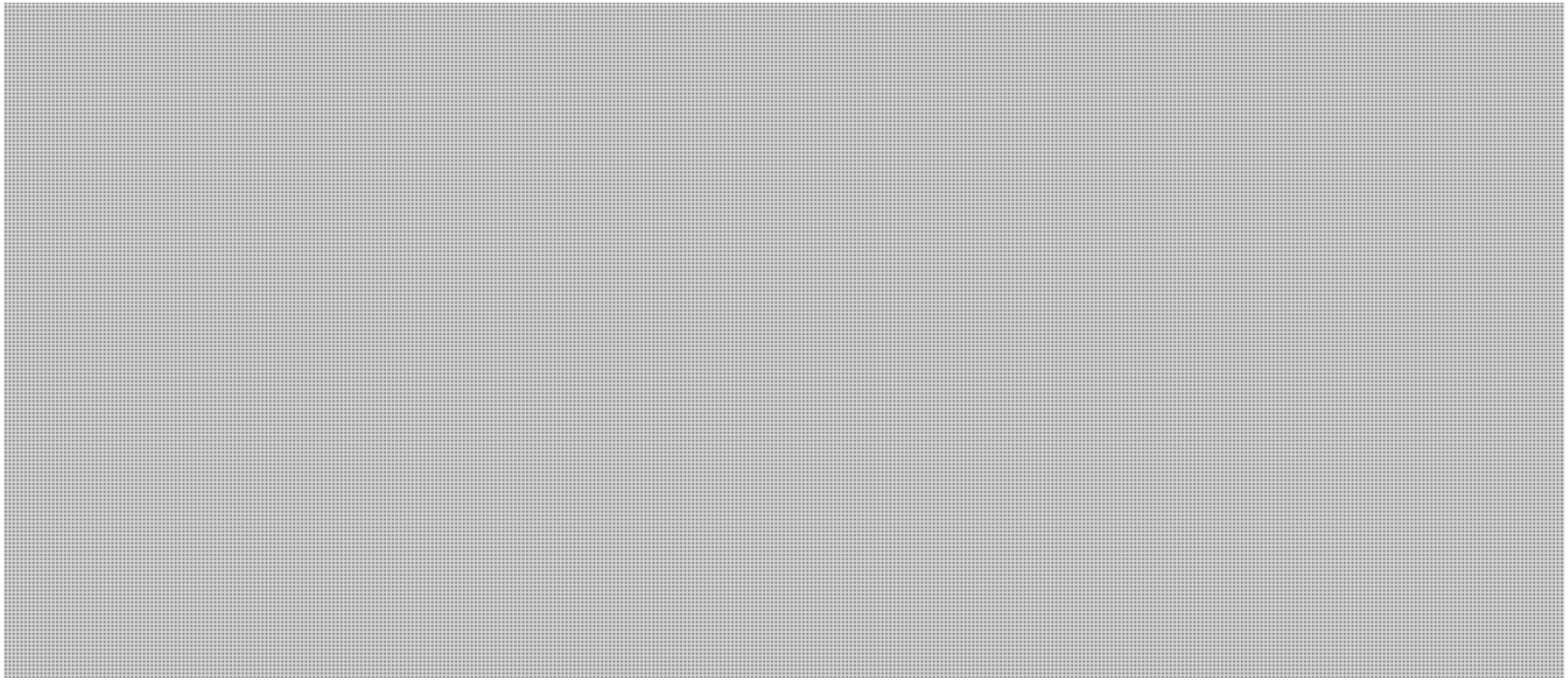
Cognitive AI Lab



Summary of Workshop

s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

Executive Summary



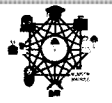
s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

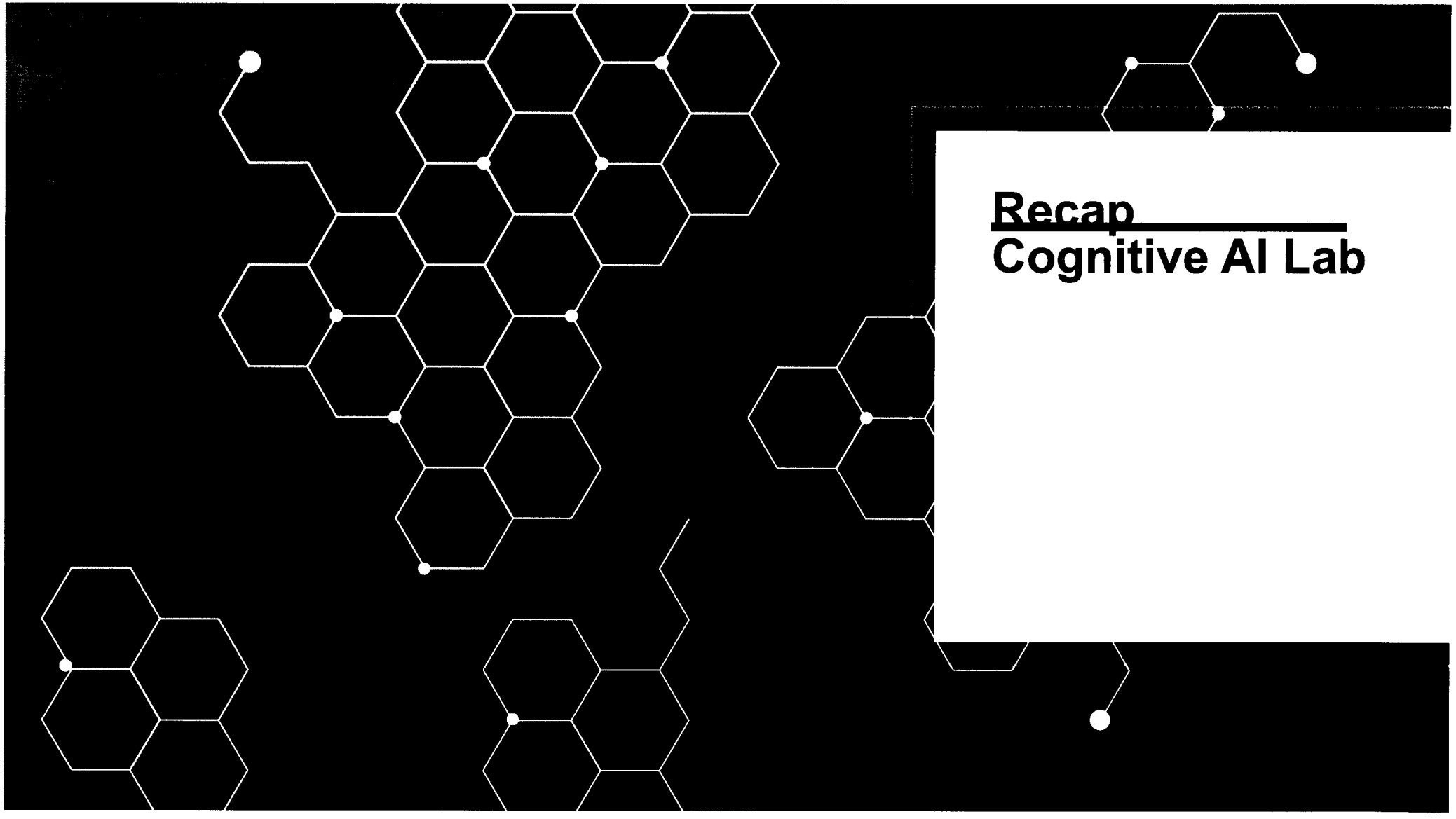
SUMMARY OF WORKSHOP

[redacted]

s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

SUMMARY OF THE LAB





Recap
Cognitive AI Lab

s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

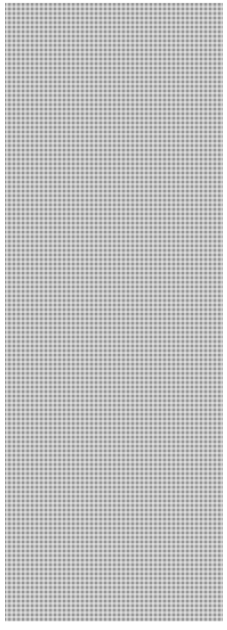
Lab Objectives

s.20(1)(b)

s.20(1)(c)

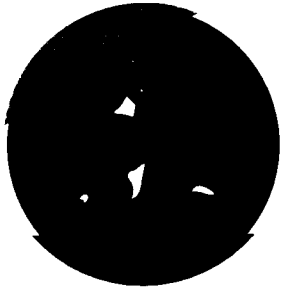
RECAP

List of Participants



- Jason Proceviat
- Gino Lechasseur
- Nancy Camacho
- Cameron Turner
- Deborah Ingraham
- Brad Sullivan
- Jeff Carr
- Anna Iglewski
- Shirley Carruthers
- Hieu Vu
- Justin Bayard
- Marc Brouillard
- Michael Karlin
- Ashley Casovan

Agenda – Day 1



9:45 – 10:00
Arrival

10:00 – 10:15
Introduction & Objectives

10:15 – 11:15
AI 101

11:15 – 12:30
Applications of AI

12:30 – 1:15
Lunch

1:15 – 2:30
Extracting Key Opportunities

2:30 – 3:00
Next Steps

s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

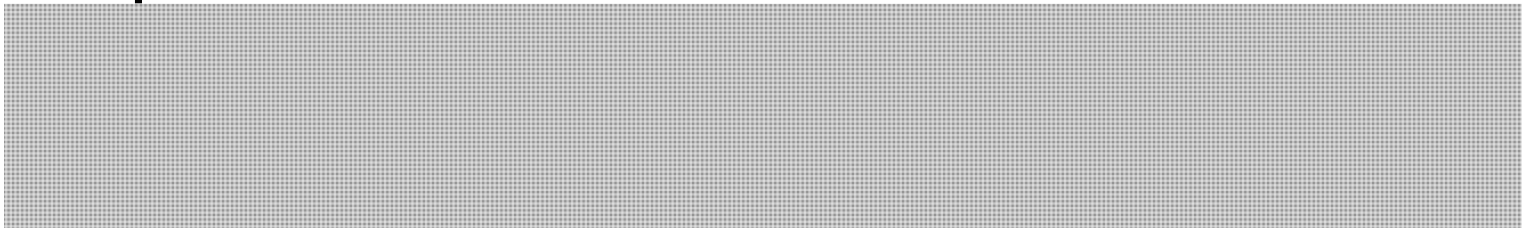
RECAP O[illegible] COGNITIVE A[illegible] LAB

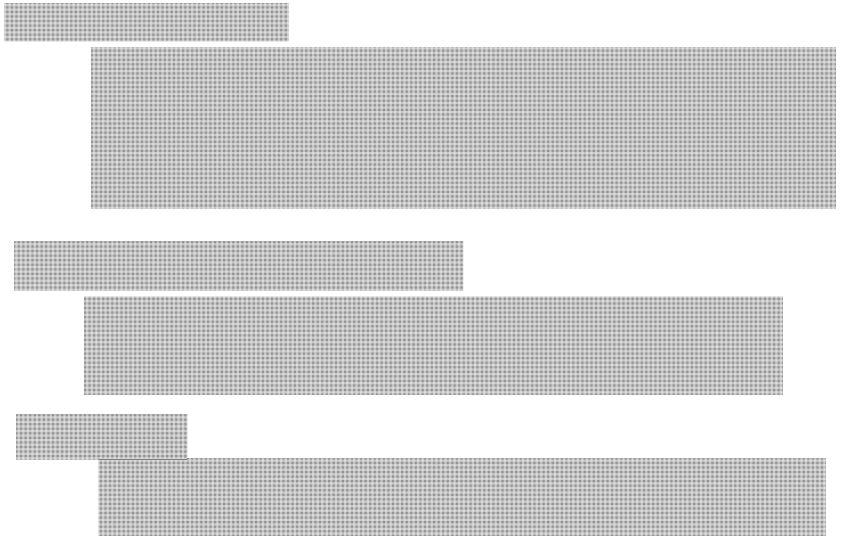
Introduction & Objectives and AI 101

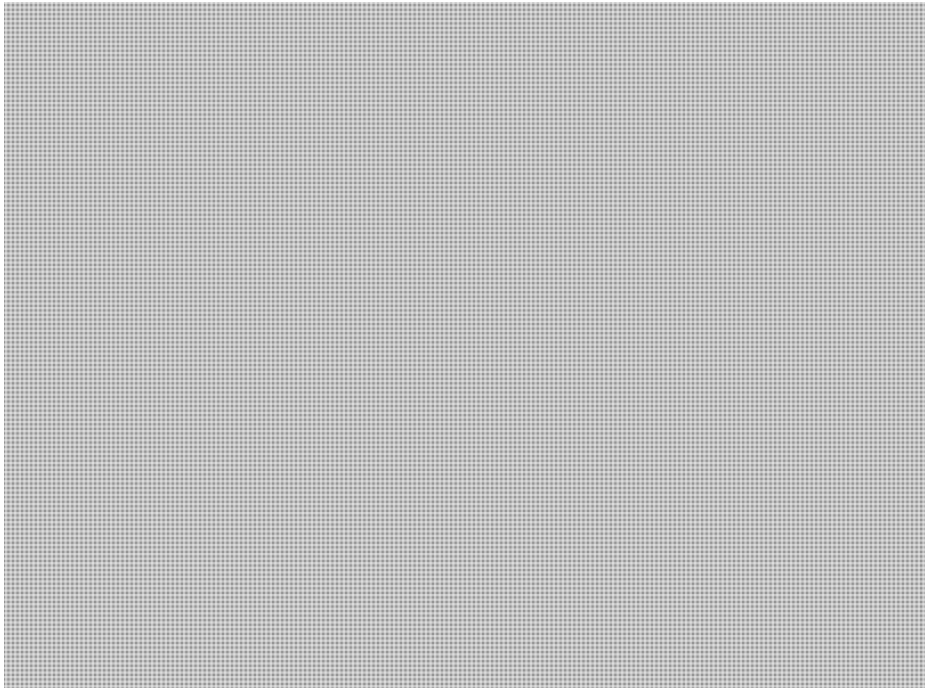
[redacted]

s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

Applications of AI, Extracting Key Opportunities and Next Steps







Agenda – Day 2

8:00 – 8:30
Arrival &
Breakfast

8:30 – 9:45
Ideate Key
Opportunities

9:45 – 10:00
Prioritize Key Opportunities

10:00 – 11:00
Deep dive into key
opportunities

11:00 – 11:15
Break

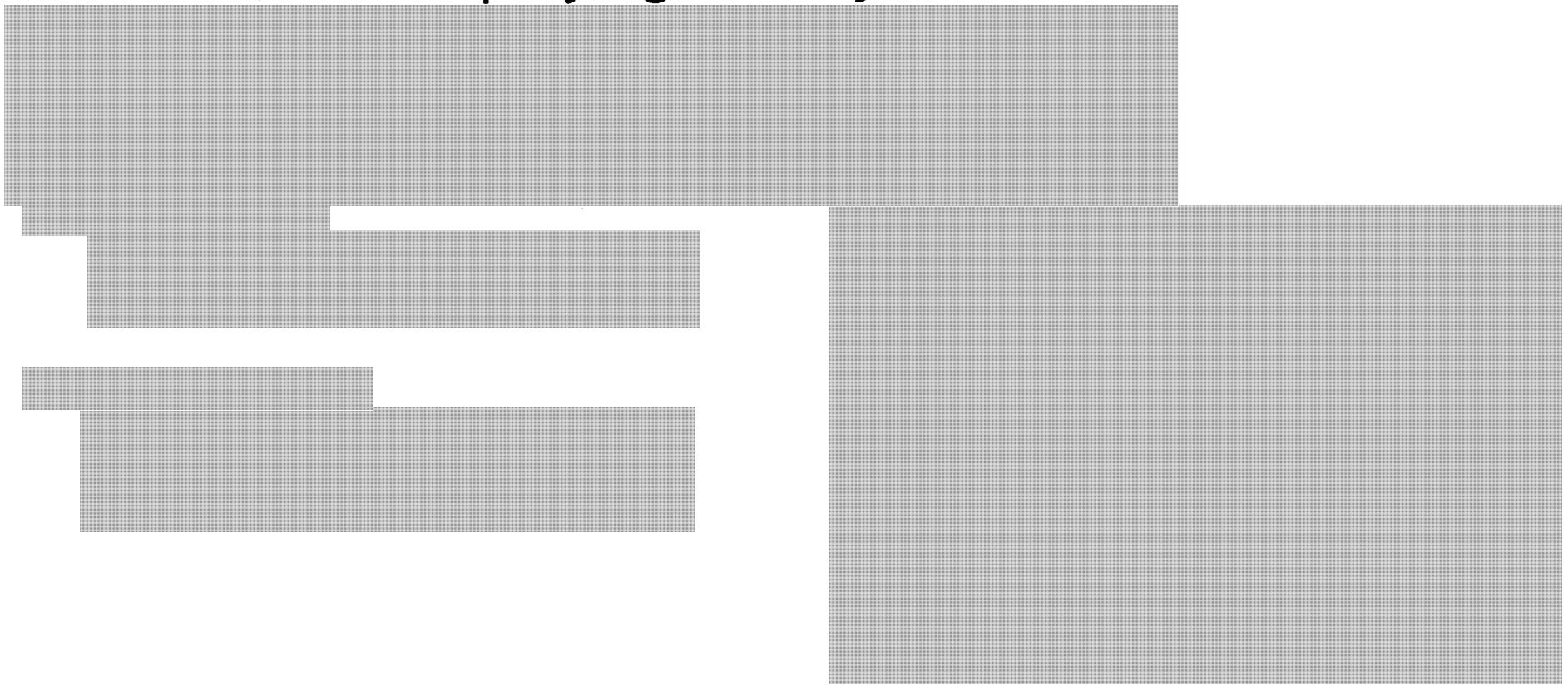
11:15 – 12:15
Define Course
Forward

12:15 – 12:30
Conclusion

s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

RECAP OF COGNITIVE AI LAB

Introduction & Deploying AI Projects



s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

RECAP OF COGNITIVE A[illegible] LAB

Concept Deep Dive

Ideate Key Opportunities

Prioritize Key Opportunities

Deep Dive into Key Opportunities

Define Course Forward

Deep Dive into Key Opportunities

- Divided into two teams within the two opportunity areas each team:
  - Laid out current state for each of the opportunity area.
  - Identified challenges and improvement areas for each opportunity area.

[redacted]

- Each team summarized their findings to the rest of the audience.

Define Course Forward & Next Steps

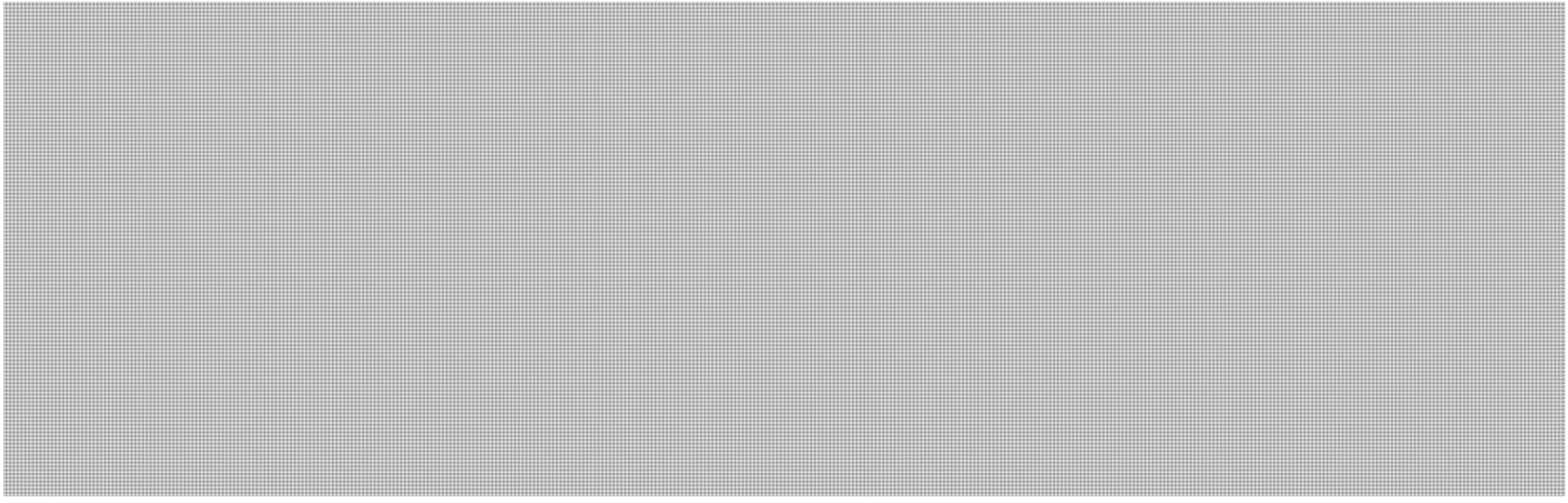
- [redacted]

[redacted]

s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

RECAP OF COGNITIVE AI LAB

What we accomplished



Appendix

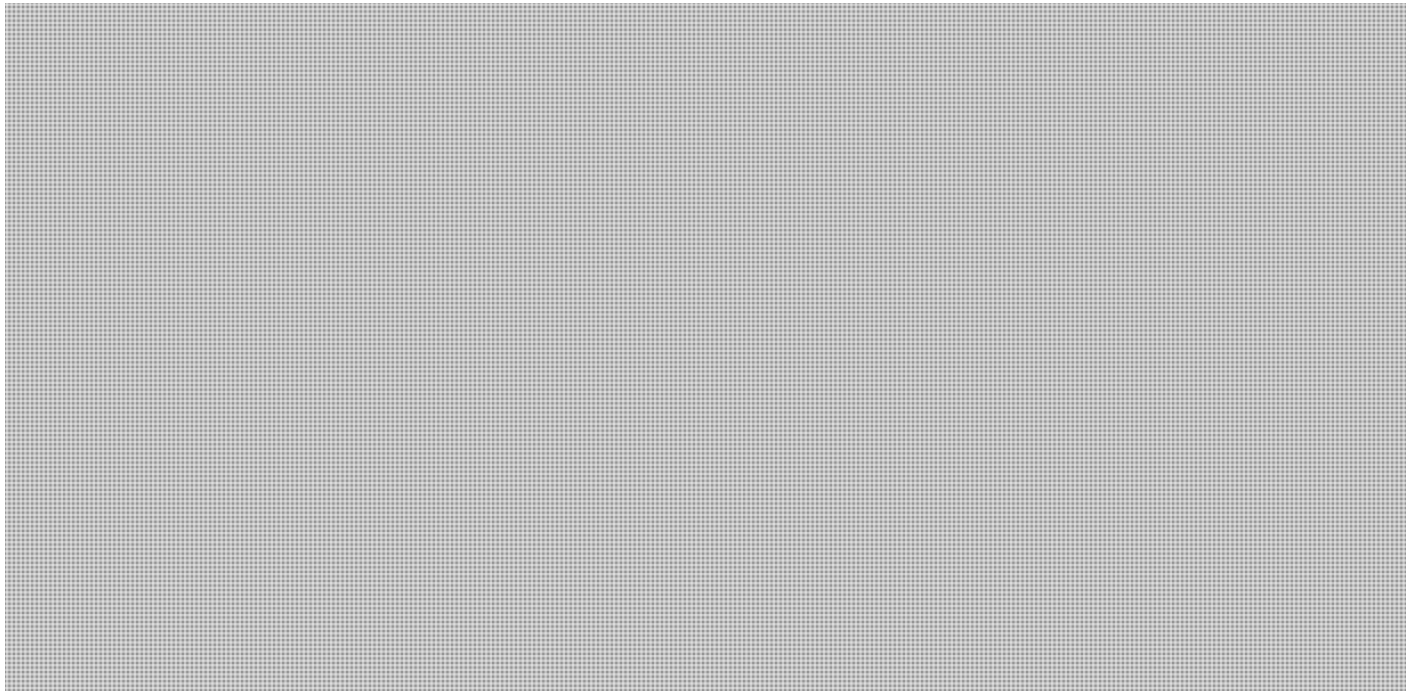
Opporunity Areas Prioritization

Future State Posters

s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

APPENDIX

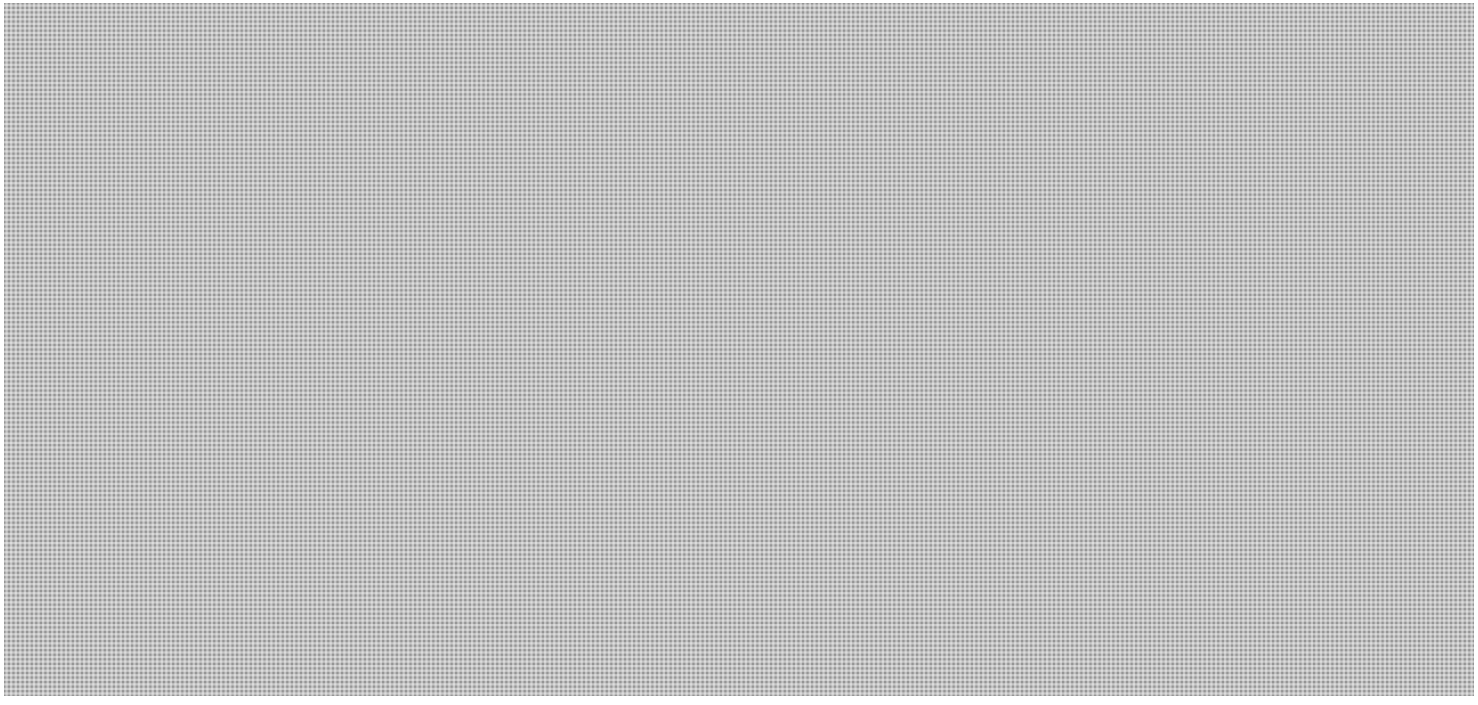
Opportunity Areas Prioritization



s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

APPENDIX

Opportunity Areas Prioritization



s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

APPENDIX

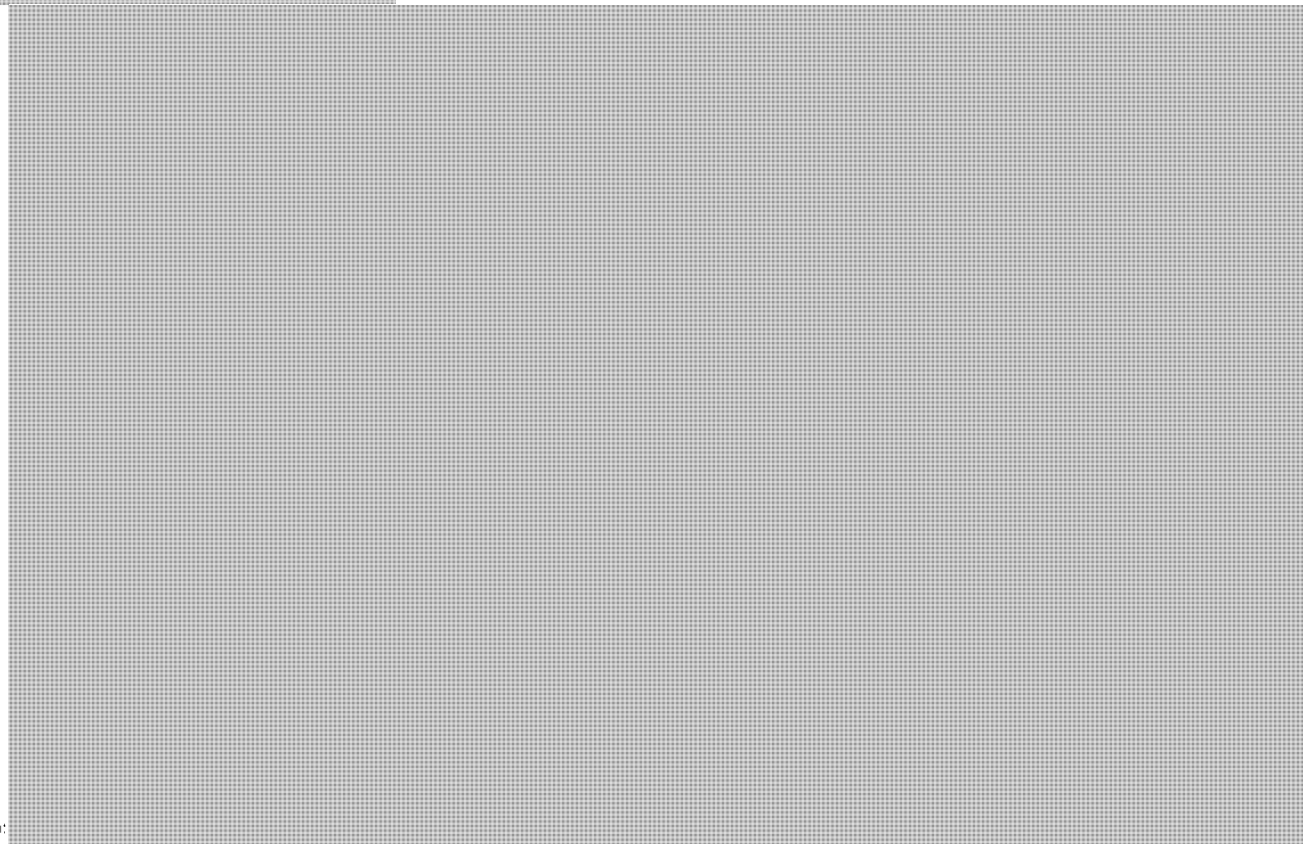
Output of Day 2: Two domain areas were dove into

s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

APPENDIX

Output of Day 2: Two domain areas were dove into

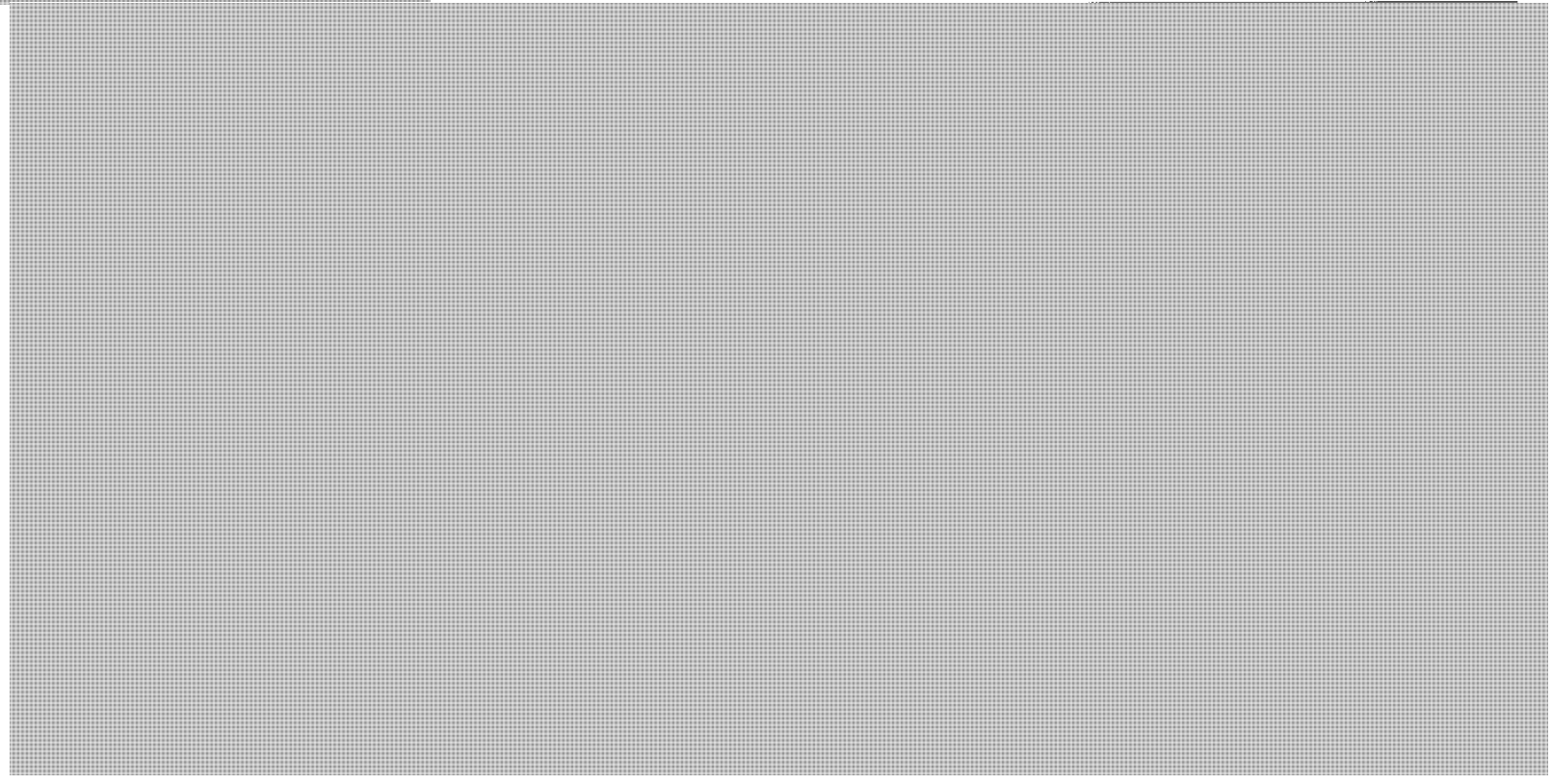
[redacted]



s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

APPENDIX

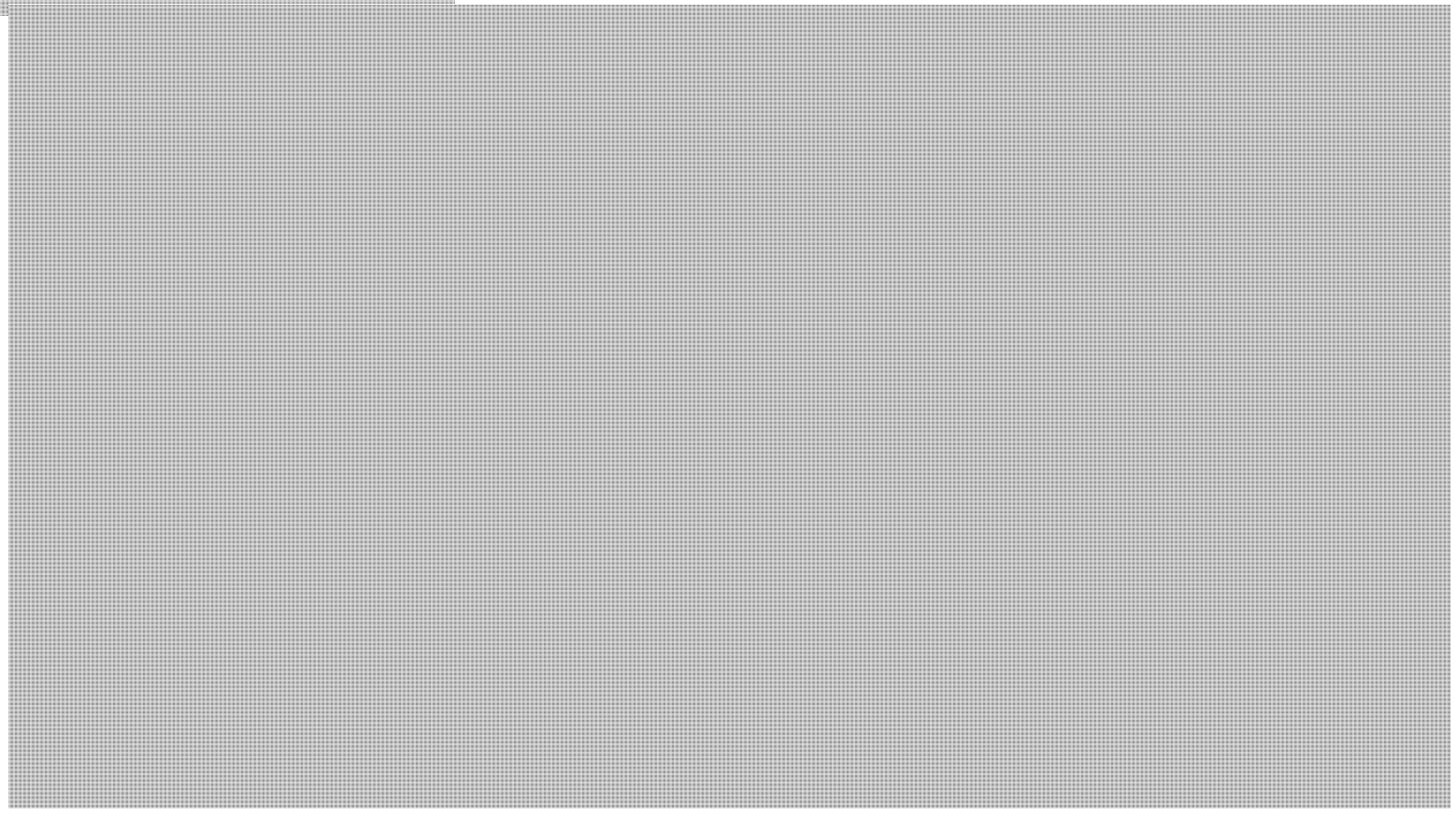
Output of Day 2: Two domain areas were dove into



s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

APPEND

Output of Day 2: Two domain areas were dove into



Karlin, Michael

From: [redacted] (CA - Ottawa) <[redacted]@deloitte.ca>
Sent: February 26, 2018 10:59 AM
To: Brouillard, Marc; Karlin, Michael
Cc: [redacted] (CA - Toronto); [redacted] (CA - Ottawa); [redacted] (CA - Kanata)
Subject: Follow-up to session

s.20(1)(b)
s.20(1)(c)
s.20(1)(d)

Hi Marc and Michael, thanks again for your sponsorship and participation in last week's meeting. We'd be interested in your thoughts on how it went and areas we might adjust if/when doing another one.

For us it was interesting and notable that a large part of the value for participants was sharing with other departments and hearing what's going on elsewhere as much as the content itself. That said, I think a richer discussion around the specific opportunities and prioritization could be had if it was only one Department in the room. So a trade-off there.

A few of the folks have requested a follow-up so we had a few questions for you:

1) Are you ok if we send a broad 'thank-you' to participants ? We can float that by you first if you like.

2) How involved do you want to be in all the follow-ups ?

[redacted]

[redacted] for a debrief on this session I'm wondering if you can help us find a couple slots in the next week or so for 1 hour in Marc's and Michael's calendars.

That ok for everyone ?

Cheers !

[redacted]

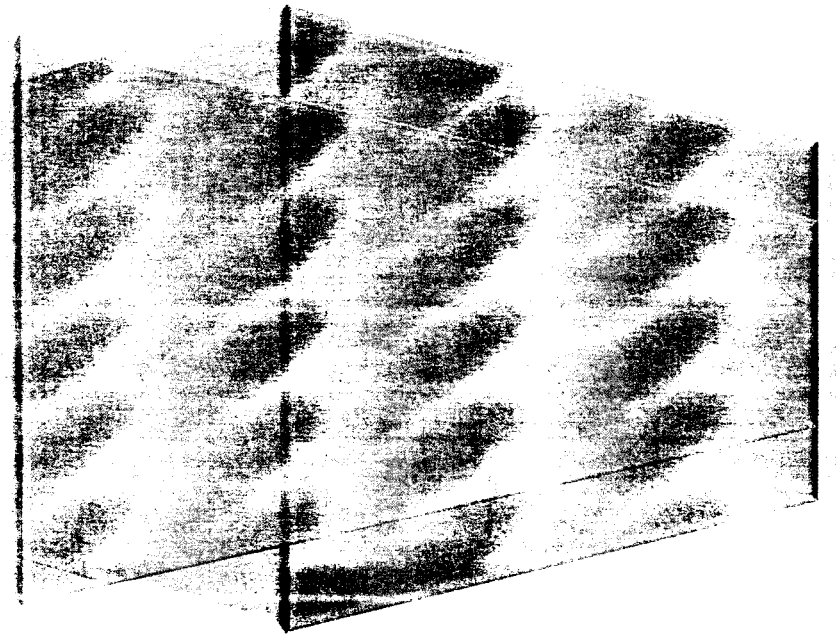
papier, la reproduction, la diffusion ou toute autre utilisation de ce message et de ses pièces jointes sont strictement interdites. Si vous n'êtes pas le destinataire prévu, veuillez en aviser immédiatement l'expéditeur en répondant à ce courriel et supprimez ce message et toutes ses pièces jointes de votre système. Merci.

Si vous ne voulez pas recevoir d'autres messages électroniques commerciaux de Deloitte à l'avenir, veuillez envoyer ce courriel à l'adresse unsubscribe@deloitte.ca

TBS-SCT

Artificial Intelligence for Digital Government

Understanding the policy, ethical and legal implications of automation



Artificial Intelligence for Digital Government

CONTENTS

VERSION HISTORY		
#	**Date**	**History**
Sections 0.1 – 0.10 – The Early Drafts		
0.1	January 26	First draft for internal consultations – TBS internal team
0.2	February 16	First round of comments included – Draft to GC DCIO
0.3	March 3	GC DCIO comments incorporated
0.4	April 7	A round of bilateral interviews was conducted with representatives of CDS, TBS Legal Services, IS&J, IPPD, and SIDM.
0.5	April 13	New section on cognitive intuition. New box on the "black box problem."
0.6	April 18	Following a meeting with Steffen Christensen and Peter Guibaut at Policy Horizons, refinements made. New section under What is AI on classification. Some semantic changes as well.
0.7	April 20	Trimmed next steps section; this is a discussion paper after all. *Draft sent to GC CIO.*
0.8	May 3	Further semantic refinements throughout. "Existing landscape" section removed so that the document reads more like a consultative piece.

EXECUTIVE SUMMARY

If a person lost their job on a Friday, what would the Government of Canada need to implement to provide them the benefits and services they need by Monday? What if the concept of business days ceased to exist? What would we need to do to eliminate service backlogs and prevent future backlogs from forming?

Artificial Intelligence (AI) is the name given to a suite of related technologies that allow computers to imitate the intelligence and reasoning capability of humans. Since the 1990s, AI applications have slowly expanded throughout the modern economy, and are now a critical element to the computing power that drives society. What makes AI different than other computer programs is its capacity to comb through and organize vast amounts of data without regular human intervention. As these applications are exposed to more information, their ability to glean insights from data improves. With time and experience, deep learning algorithms can gain a level of analytical efficiency and accuracy that far outstrips human capacity.

These applications have the potential to significantly change how government works, and specifically, how it offers services to its residents and businesses. Over the past few years, developers have made extraordinary progress in developing AI that can communicate with clients directly in fluid, natural language like a front-line service agent. AIs exist that can process the eligibility of an applicant for a service or benefit in seconds. AI can even generate sensible, grammatically-correct, and relevant textual summaries, like those being used extensively by newswires. The efficiencies provided by AI could be one of the most radical steps forward in service delivery since the invention of the Internet.

These technologies are not without risk, however, and so if federal institutions are to proceed with deploying AI applications to improve service delivery, they will need to take several policy, ethical, legal, and human resources issues into consideration. For example, how can we encode AI to make moral decisions? Can an AI make an eligibility decision by itself, and if so, how does that affect Ministerial accountability? How will widespread automation affect the workforce of the public service? All federal institutions considering this technology will likely grapple with these and other questions.

The objective of this discussion paper, which is intended for layperson policy managers, is to identify the key questions to be addressed when deploying this technology, as well as lay out a path for consultations. While automation is a powerful economic trend that governments worldwide must face, many of these issues fall outside the mandate of TBS. As such, this paper focuses specifically on three key applications of AI for digital services; these applications have similar issues related to them.

- Virtual front-line service agents
- Processing eligibility of benefits or services
- Intelligent analytical tools

INTRODUCTION

In February 2011, an IBM-built computer by the name of Watson was victorious in three matches of Jeopardy against two of the most successful champions since the game's inception. Despite computers beating humans in high-profile games such as the famous 1998 chess match between Garry Kasparov and Deep Blue, Watson's victory was different. Unlike chess, winning Jeopardy required the computer to understand natural language, including all of its nuance and symbolism. It had to evaluate puns, rhymes, and abstractions. In short, it had to think more

laterally than computers had ever done before. Watson won three games of three, and acted as a very public herald for the beginning of the age of artificial intelligence (AI).

While this technology was born in 1956, over the past decade, AI applications have been deployed in such variable and extensive ways that it increasingly drives the modern economy, including in sensitive areas. AI has replaced humans on stock market floors[1] and the management of multi-billion dollar hedge funds.[2] It assists with medical diagnoses and operates complex machinery autonomously. Corporations worldwide are examining workflow automation to increase efficiency of their operations. AI agents are beginning to understand, and speak in, natural language to interact with humans using sophisticated intelligent virtual agents. They interact with humans patiently and responsively, and have shown to effectively assist with people with difficult tasks such as learning languages.[3]

Anthropomorphizing software

When faced with a few lines of text, an AI can be taught to understand meaning from them, but it is an entirely different cognitive process than what goes on in a human brain. As it is informed by more data it improves, a process that mimics learning but is not the same process as human learning. This paper will speak about AI using humanlike semantics from time to time because it is a helpful way to communicate technical concepts, but it is important to remember that fundamentally, AI is software, not a conscious being. For now.

There is a very high likelihood that by 2025, AI will touch every aspect of modern society in ways both visible and invisible to Canadians.[4]

Canada is positioned to emerge as a global leader in AI research, development, and application. Budget 2017 committed $125 million to launch a Pan-Canadian Artificial Intelligence Strategy to support these clusters attract the talent they need to maintain their advantage.

The Government of Canada is looking into how it can harness the opportunities provided by AI to provide more timely services to citizens and other clients[5], as well as improve efficiencies in its operations. Federal institutions are working towards offering better user interfaces to make their services easier to use, but these gains will not accomplish a frictionless service environment if the person faces weeks-long backlogs in having a benefit application processed. AI systems can work faster than human equivalents, and will work over evenings, weekends, and statutory holidays. They do not tire or suffer as much impairment that can affect their judgement. AI systems can be deployed by service departments and agencies to answer questions posed by clients – as well as made eligibility determinations – in order to vastly improve the efficiency of service and virtually eliminate backlogs.

---

[1] http://www.bbc.com/news/business-34264380

[2] https://www.theguardian.com/technology/2016/dec/22/bridgewater-associates-ai-artificial-intelligence-management

[3] Macedonia, Groher and Roithmayr, "Intelligent virtual agents as language trainers facilitate multilingualism." Frontiers in Psychology. April 2014. Link: http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00295/full

[4] A qualitative survey by the Pew Research Center of over 2,500 academics, policy analysts and corporate executives found broad consensus to support this prediction. While the study was American, respondents were international. See: Pew Research Center, "AI, Robotics and the Future of Jobs." Link: http://www.pewinternet.org/files/2014/08/Future-of-AI-Robotics-and-Jobs.pdf

[5] This paper uses the term "clients" to represent the diverse clientele that use Government of Canada services including, but not limited to, citizens, permanent and temporary residents, and businesses.

On the other hand, relying on a complex AI system to make decisions is potentially risky, both that the system fails, but also that the decisions that are being made are actually furthering the goals provided to them by officials. How can we know that an AI is interpreting data in a manner that is responsible? How do we teach it social or geographical context such that it can make decisions in a nuanced fashion? What are the workforce requirements in a post-AI world?

Governments worldwide are now grappling with the consequences of a technological development that is transforming service delivery across sectors. The United States, United Kingdom, France, and Japan have undertaken high-level examinations of AI applications within their respective governments and on their economies writ-large. The Government of Canada has the opportunity to build on the brain trust of private sector and academic leaders in this field to position itself as a world leader in AI for digital governance.

While AI is undergoing rapid advancement, it is important that the policy, ethical and legal implications of the use of this technology to deliver government services be addressed methodically. The service delivery opportunities are significant, as are the pitfalls.

PURPOSE AND SCOPE

Artificial intelligence is and will be used across many sectors of the economy in ways that the Government may take interest; however, the scope of this paper is limited to its provision of services or benefits to the public. This includes the automation of the administrative decision process or how departments or agencies interact with their clients. With the approval of the *Treasury Board Policy on Service* in 2014, TBS is working with service departments and agencies on means to ensure that services are client-centric and e-enabled; judicious deployment of AI can contribute to these goals. There are additional uses for artificial intelligence in a Government of Canada context such as internal services automation or policy impact simulation, which may fall under other policy instruments.

This discussion paper will examine the policy, ethical and legal considerations around the use of this technology within the Government of Canada. It should not be construed as an implementation plan, feasibility study of available technology, or a business case. Use of a hypothetical example as how AI may be deployed within a Government context should not be construed as a plan to do so.

Throughout the paper, illustrative examples are used to show how this technology can be beneficial to governance. These vignettes – prominently located in green boxes – do not represent any existing plans of the Government of Canada and should be considered theoretical only.

This is a consultative paper; questions are asked throughout the document for you, the reviewer, to provide your opinion. Your feedback and responses may be used to inform the development of guidance in the future.

INTENDED AUDIENCE

The primary audience for this paper are non-expert policy managers that would be responsible for the design, implementation and/or auditing of an AI system within their respective department, agency, or Crown Corporation.[6]

[6] "Institution" refers to federal departments, agencies, and crown corporations.

WHAT IS ARTIFICIAL INTELLIGENCE?

According to John McCarthy, the computer scientist who invented the term artificial intelligence (AI) in 1956, its definition is "the science and engineering of making intelligent machines." As technology has evolved, AI has grown to become a broad term that generally refers to a computer program that can simulate human cognition. It attempts to marry the speed, predictability, and rigour of computer processing with the lateral and abstract thinking that the human brain is capable of.

PATTERNS AND PREDICTIONS

A key feature of AI is its ability to identify patterns in information and make predictions based on these patterns.

Humans are adept at identifying features that are common to several processes or events. We provide names to these patterns and understand them as if they were real, but they aren't. This process is called "abstraction." When we speak about "politics" or "economics," we are speaking about a collection of ideas, institutions, and objects that are related to one another. So we know, for example, that currency is more relevant to the concept of economy than "dust" or "pizza." To humans this ability is innate, but computers have difficulty understanding abstraction without being taught. AI allows a computer to begin to connect concepts like a human can, which it can use to make predictions as to how – and how much - concepts interrelate.

Using its pattern recognition ability, AI has been programmed to understand natural language, spoken and textual. It has learned how to perceive the world through machine vision, or understand complex systems enough to make predictions and optimizations similar to a human.

MACHINE LEARNING

AI can be designed to get better at its tasks based on the amount of input information that it has access to. As AI accesses more data and identifies more relationships between data, it is able to improve its performance of whatever analysis it has been tasked to do. That said, machines do not "learn" like people do; machine learning is actually a process of becoming iteratively more precise and accurate with how it sorts through information.

This iterative improvement can be done without the direct interference of a human, and this will take an AI far, but sometimes human interjection is necessary. For example, an AI tasked with sorting through imageswill quickly learn the difference between an image of a duck and an image of a pie, but it would take significantly longer to recognize the difference between an image of a duck and an image of a duck-shaped hunting decoy.

Machine learning is used extensively throughout many sectors, from Amazon.com recommending what books you might like to read, to smart thermostats figuring out the likely times their users are home or away.

AUTONOMY

The Pitfalls of Precision

In his book Superintelligence, *Oxford University professor Nick Bostrom warns that the boundary between AI and general computing is fuzzy, and that one can't find exact precision in a definition to satisfy all needs. As a result, this paper takes a broad approach to defining AI, roughly in line with the IEEE's definition.*

The AI applications that most of us know are intelligent autonomous agents; that is, they are capable of operating on their own following given instructions towards a specific goal without regular user input. For example, an online banking interface that allows the transfer of funds from one account to another is not AI, but a program that watches transactions for suspicious activity, alerting security personnel when there is a potential issue. The former is inert without direct user input, whereas the latter can run indefinitely along the instructions that it has been given. "Autonomy" in this sense does not mean that systems should not go unsupervised, it just means that it does not need regular user input.

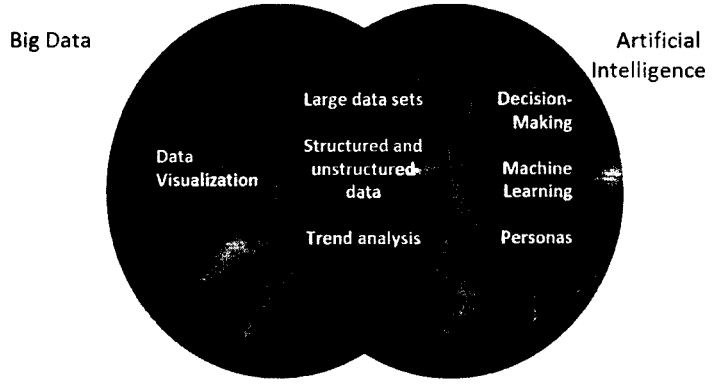
AT HOME IN LOOSE STRUCTURE

Traditionally, the efficacy of a computer's ability to analyze large datasets was proportional to how much structure there was in the data. "Structure" refers to the limitations on possible answers, about how much complexity the computer needs to make sense of. Like a date of birth or last year's income, a structured dataset has a clear limitation on the answers that can be provided. A response to "what is your date of birth" can't be "a car" because it does not fit within the structure of the data.

Services can be complicated, however, and sometimes departments need to collect – and understand – unstructured data. A good example is an Employment Insurance application, which may require that the applicant provide a detailed reasoning as to why they are unemployed. These reasons vary significantly in content or the language with which they are expressed. Unstructured data used to be notoriously difficult for computers to glean insights from, but advanced AI systems are now capable of sorting through information quickly and building structure in the information itself. With enough data, and using abstraction, an AI system can be programmed to group terminology (e.g. "let go"; "professional disagreement"; "different direction" = terminated from position). It can then analyze and glean insights from this information as if it was provided to the system in a structured fashion.

This ability is not unlimited. AI analysis can suffer when it is not provided any structure at all, or when the potential output is infinite. Even unstructured sections of EI applications can have expected results, with descriptions of terminations from employment being limited. With some services however – such as a patent application – the potential input information is near-infinite. It is also why the conversational skills of an AI are limited; ideas can be expressed in a number of ways and the AI needs the ability to detect subtext or inference from the limited data provided to it.

AI AND BIG DATA ANALYTICS



For the past several years, Big Data has been another tool on the frontier of governance. At first glance, Big Data is similar to AI; both involve trying to make sense of vast data sets. Both involve gleaning insights from patterns that may be too difficult for a human to identify on their own. So what is the key difference?

If AI is a classification and pattern recognition engine, then big data is the fuel. Characteristic of almost all AI systems is the need to train using data and experience.

The concepts overlap considerably; AI simply adds a layer of learning, prediction, and decision-making. Its ability to improve itself without direct human intervention is its defining feature. As well, AIs used as virtual agents (see below) often have personas which they use to interact with clients. Conversely, Big Data analytics tools often use a form of visualization to produce graphs or graphics that help users understand the data that is being presented to them. This feature does not require autonomous intelligence to perform.

As research and development in this field expands significantly and deployed into real-world applications, it is easy to be overwhelmed by the terminology. "Big data," "deep learning," "neural networking" and other terms all describe related concepts. At its core, these terms all describe ways that machines learn, comprehend, and use data.

WEAK AI VS STRONG AI

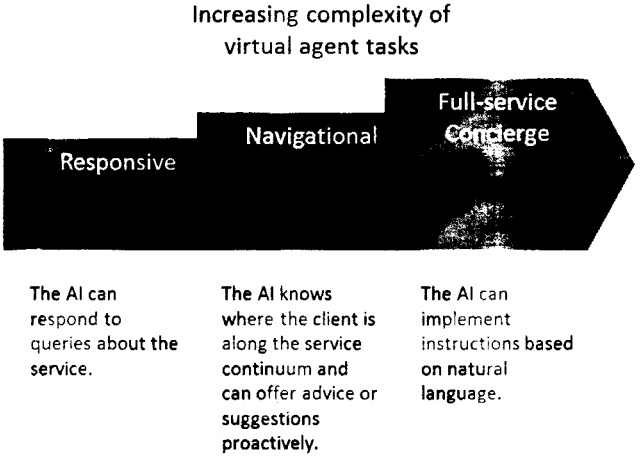
From playing chess to predicting the weather, in contemporary times, computer programs tend to be better than humans at lots of things. The limitation behind this trend is that a single computer program tends to only be good at one specific task. Whereas a human frontline service agent can also learn to knit, cook, or translate a language to mastery, Microsoft Excel can't learn to be a flight simulator. So computers are – at least today - superior to humans in *specific* intelligence, but fail to match humans' *general* intelligence.

The applications of AI discussed in this paper are generally known in the field as "weak," because while they bring an unprecedented sophistication to the performance of their niche tasks, they will never be as polyvalent as human workers. Researchers are working steadfastly towards "strong" AI, or computer programs capable of developing as much (or more) expertise in a wide variety of tasks. While this development would bring extraordinary social change, its development is at this point nascent, and falls outside the scope of this paper.

AI AND APPLICATIONS IN E-GOVERNMENT

Self-learning, interactive computing offers extraordinary promise to e-government initiatives, and governments around the world are just starting to understand its full potential. How can the Government of Canada use AI to evolve its service delivery to clients and efficiency of its operations?

This paper focuses on three applications for improving service delivery to clients and improving efficiency of operations. All of these applications exist and are functional; virtual agents and eligibility processing are used across various sectors, whereas AI process automation is in the early stages of implementation.



VIRTUAL SERVICE AGENTS

End-to-end digital services are the frontier of governance in the 21st century; that is, the ability to access the entire continuum of the service from application to delivery without the need for a paper form, or for the user to have to interact with a service agent. Ideally, the service experience from authentication to application to receipt of benefit or issuance of payment should be a seamless process that does not require the use of a phone or visit to a

service centre.

For the client, these other channels are inherently less convenient; opening times are restricted, require waiting on hold, or involve travel times. For some business owners, especially small businesses, lengthy wait times, or the requirement to access services during business hours can lead to an unacceptable loss in productivity. For the government, digital transactions are considerably less expensive. Assuming that the digital service offering is understandable, convenient, and accessible enough for someone to want to use it, there is incentive for all parties of a service transaction to want to move to the digital channel.[7]

Reality is more complex than theory. People have complexities in their lives that complicate a service experience, or have difficulty understanding the language used. In these circumstances, people use a variety of tools to help them find the information they are looking for, whether frequently asked questions (FAQs), social media, or personal networks.[8] If these do not work or are otherwise inconvenient (i.e. there is a time pressure), a client will likely turn to another channel to speak with a human service assistant.

A conversational AI interacting directly with a client can play a role in keeping them on the digital channel. These virtual agents[9] are capable of parsing natural language into searchable terms, accessing information located in FAQs, manuals or even specifically-identified internal documents and reply to the question in a way the client can understand.[10] With additional information and user feedback, the virtual agent will continuously improve at this task without the need for direct human intervention.

The capabilities of an AI virtual agent can be scaled up over time to provide expansive levels of client care as it gains experience and improves the way it manages information. It can offer responsive services, answering queries related to services passively. Eventually it can expand to become more navigational, offering hints, advice, or step-by-step instructions more reflexive of where a person is in the continuum of their service experience. Eventually an AI can be capable of actually executing instructions, such as accessing and pre-filling a form based on natural language.

The Canadian Imperial Bank of Commerce (CIBC) deployed a virtual agent for customer service in 2008. It experienced a 25% deflection in calls to its call centre and 50% reduction in email volume.

([24]7 and CIBC, Accessed Jan 2017)

Virtual agents help filter routine questions away from human service agents so that they may focus on helping clients through complex or distressing cases, or cases where a client is uncomfortable relaying their circumstances to a machine. This technology has been deployed successfully in the private sector, and is being trialled in a government context in the United Kingdom. The county council of Enfield in North London will debut the use of

[7] Understanding that, for the foreseeable future, there will be individuals or businesses that prefer to access services through other channels.

[8] Leggett, Kate (Forrester), Trends 2016: The Future of Customer Service. January 2016. See summary of findings here: http://blogs.forrester.com/kate_leggett/16-01-28-online_self_service_dominates_yet_again_why_its_an_effortless_way_to_get_to_your_answers

[9] There are a variety of names given to programs with similar functions such as, but not limited to, intelligent personal assistants (IPAs), virtual customer service agents, virtual agents, chatterbots or chatbots. Nomenclature may vary in cited sources.

[10] All of this capability is in text form only; virtual agents for voice interaction are not advanced enough to consider in this context,

IPSoft's Amelia virtual agent program to assist with residents wishing to use its digital services.[11] Its expressed objective is to reduce transactional queries regarding its services so that its service agents can focus on complex cases.

This technology has advanced significantly over the past five years and is expected to continue its rapid advancement for the next decade. Unfortunately it does have limitations. As described above, conversations carry a lot of information outside of the basic text. Emotional queues or the use of sarcasm and humour can quickly confuse an AI conversational agent. While they are adept at managing basic questions, a lengthy, interactive conversation is not possible at this time. Some virtual agents, like Botler, provide a user with a defined set of potential inputs to reduce errors in the conversation, which results in a more scripted interaction.

An additional benefit of virtual agents is their ability to structure through a standardized approach to collection. Through interactions with clients, a virtual agent can help reduce spelling errors, inappropriate entry of dates and addresses, etc. This improves overall data quality, which in turn could help eligibility determination.

How might it work? Virtual Agents*

After over 25 years of owning the little florist on the corner, Geneviève has decided to retire and sell her business. Between the transfer of the business and her CPP/OAS applications, Geneviève is facing a paperwork burden. At the behest of her daughter, she logs on to the federal single online window for government services, but doesn't know where to start. She sees a little bubble at the bottom right-hand corner that says "Click me for assistance." She clicks the bubble.

A virtual agent pops up, introduces itself, and asks why she has logged on. Rather than state the services she needs, Geneviève explains her life situation; that she is retiring and wants to sell her business to her assistant manager. The virtual agent is able to parse out her needs, organize them in order if required, and starts to walk her through the service process. "No problem," it replies naturally, "Let's start with the business transfer."

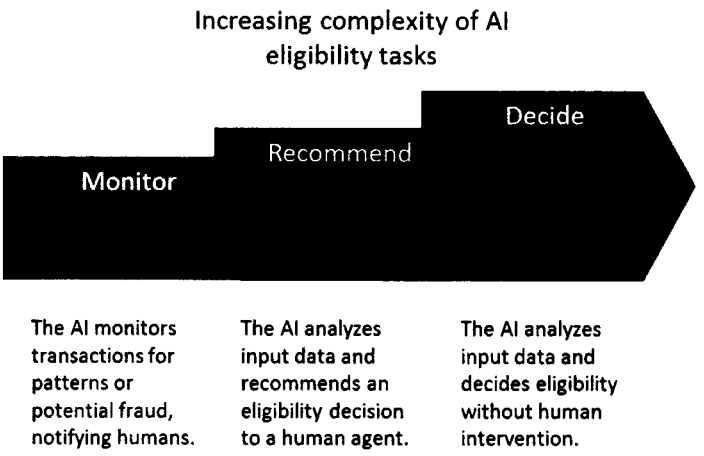
The virtual agent remains open, providing advice along the way as to what to fill in each field, walking Geneviève through the process.

This example, as well as the others below, is included for illustrative purposes only. It shows a possible future state, not a planned activity.

ELIGIBILITY PROCESSING

Designing services well so that clients want to use them is important, but its benefits are lost if the wait time to receive eligibility decisions on services is too long. Part of serviced excellence is cutting wait times, and AI can play a role here as well.

[11] http://www.forbes.com/sites/jenniferhicks/2016/06/28/artificial-intelligence-helps-uk-council-handle-community-inquiries/#55dc0152719d



Processing service applications requires that an analyst review application information, verify to see if it is true and believable, and checking if the information that has been submitted meets the program's eligibility criteria. This process can take time, both due to the amount of information collected as well as the limitations on resources. By using all appropriate, program-related input data; pattern recognition; and machine learning, an AI can process eligibility decisions faster than a human. This application of AI would allow eligibility analysis to be processed outside of core work hours, for data analytics to be gleaned and acted upon promptly and organically, and for patterns to be established so that particularly complex or unexpected applications can be investigated more thoroughly.

This level of decision automation has been tested and deployed in private sector settings for over a decade. Insurance and financial sectors have been pioneers in decision automation to improve service response times and to increase fraud detection. These sectors have similar challenges to governments: mission-critical systems with many dependencies, limited budgets and competing priorities for IT development, and a desire to maximize transaction throughput and minimize fraud.[12]

Many government services have existed for decades, so there is a significant volume of potential training sets to teach AI how to process eligibility. By showing AI examples of successful versus unsuccessful applications, it can determine the necessary patterns to extend this reasoning to a new application on its own, effectively mimicking the experience of a human.

[12] See McKinsey report, "Automating the bank's back office," Link: http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/automating-the-banks-back-office

How might it work? Eligibility Processing

Jonathan has applied to renew his passport using the new online tool, where he can reconfirm his personal information and submit his updated digital picture. Satisfied, Jonathan leaves his computer and goes to brew a cup of tea. His information is received by an eligibility processing AI, who analyzes the application, noticing:

- Neither his contact information nor his physical features have changed since last renewal,
- Facial recognition software (currently in use) validates that the picture matches his previous application,
- His entry/exit information notifies that he has not travelled anywhere high-risk,
- He has changed jobs twice since his last renewal, and
- There are no flags on his file from security agencies.

The AI determines that he is eligible for a passport renewal and recommends to a passport officer to approve, with a short summary of the reasons behind the recommendation. Satisfied with the recommendation, the officer orders the creation of the new passport. The total time of eligibility determination has been reduced to minutes. As a result, there are no backlogs; Jonathan can order his passport on a Monday and receive it by Friday without having to pay for a special level of benefit.

This illustrative example makes a number of assumptions; it assumes information sharing authorities exist and system interoperability with all necessary partners is in place.

INTELLIGENT ANALYTICAL TOOLS

In late 2014, the Associated Press newswire, partnering with AI developer Automated Insights, automated its reporting of business' quarterly earnings.[13] The system, estimated to be able to generate 2,000 news articles a second, summarizes corporate earnings reports, and converts them to rote stories posted on the wire. After several months of training, configuration, and maintenance, the system is now able to post stories without any human intervention at all.

The AI journalist is capable of doing this because a) there are thousands of examples annually and the result is a dataset large enough for the computer to extract best practices, and b) most of these reports contain only factual information, with limited nuance. This has freed up journalists' time for other, more detail-oriented work.

This technology can likely be adapted to a number of government documents that are produced on a regular basis in large quantities that are often factual and follow a certain formula or template. While certainly incapable of making normative considerations and analysis, this technology can be useful to summarize and compare. For example, it would be able to write Ministerial correspondence, background sections of briefing or meeting scenario notes, background of Question Period notes, etc. This would allow human public servants to focus on analysis, policy lenses, considerations, and strategies for next steps.

[13] The Verge, January 29, 2015. Link: http://www.theverge.com/2015/1/29/7939067/ap-journalism-automation-robots-financial-reporting

How might it work? Intelligent Analytical Tools

Hamid works for Transport Canada, which is investigating whether to construct a new interprovincial bridge. Through a public consultation, TC has received over a thousand emails from a diverse group a respondents – from large stakeholders to individual, nearby residents. Others have provided comments on GCCollab, Twitter and on the Government of Canada Facebook page.

With a touch of a button, Hamid uses his GC Insights AI to:

1) **Quickly scan the letters and publically-available comments, and**
2) **Draft portions of a report summarizing these insights, grouped into categories of respondents.**

GC Insights is capable of producing these products in moments. Hamid focuses his time on applying intellectual value: analysis, providing options on how to proceed, language and nuance. This system has dramatically improved Hamid's productivity, taking a task that would have taken hours of reading, analyzing and drafting into one hour of higher-value work.

UNDERSTANDING THE IMPLICATIONS

The potential service improvements and benefits to productivity that can be realized through the implementation of AI are significant; however the introduction of autonomous systems brings with it a series of ethical, legal, and other issues that must be addressed so that benefits far outweigh risks.

While this section poses a number of questions around these key issues that will take time and wide consultation to answer, ultimately these questions can be combined to ask: How much of the decision-making process can or should be automated?

POLICY AND ETHICAL ISSUES

While the use of AI offers a lot of promise in improving the efficiency of government services, it is important to approach its use with a strong ethical foundation. Machine ethics have been debated for years, and the Government of Canada could learn from these ground breaking discussions to ensure that this transformative technology best serves the interest of the people it serves.

As these agents grow to operate in increasingly sophisticated spaces, they may act on behalf of the Crown; should they be subject to similar values and ethics as public servants?

GUIDING PRINCIPLES

The widespread deployment of AI in government services could have a significant impact on how the government provides services to Canadians, and so establishing an ethical baseline for how to proceed – a series of guiding principles to which policy managers should adhere – seems to be a necessary step. Coming to a consensus on these principles poses several challenging questions.

- To what ethical standard should we train an autonomous, intelligent computer system?
- How do we reach a consensus on the ethical principles of how to proceed, and at what level do these principles require approval? To what extent should Canadians be involved in this discussion?
- Should conformity to international norms and standards on ethics for autonomous computing, such as the Institute of Electrical and Electronics Engineers' *Ethically Aligned Design*, be central to our approach or should we focus on a made-in-Canada solution?[14]

RISK

Automated vehicles have driven for hundreds of thousands of kilometres with few incidents, illustrating at how AI is capable of making split-second decisions that can affect the health and safety of individuals. IBM has deployed its Watson AI in several major hospitals in the United States to assist with medical diagnostics. This challenges the perception that higher-risk functions cannot be entrusted to computers. The question that remains is not *whether* an AI can handle high risk, but whether it *should*.

"Risk" is a broad term. It could mean risk to an individual – whether health, financial, or otherwise – borne from a slowly-processed application, to the risk to the state borne by an AI incorrectly approving an application to a sensitive program by an individual who poses a threat.

Consider that an AI eligibility processor might flag a transaction as risky simply because it is a statistical outlier. An individual with a rare disability may have a particular set of needs that – on a population level – may not typically be combined. As well, consider certain high risk applications such as air traffic control, access to an airport tarmac or security clearances. While part or all of these services may be automated, they each bring a high risk to national security.

- Are there services that require a human decision due to risk?
- If appropriate, are there existing means to measure to rank risk that may leveraged or does a new risk test need to be designed?

PREVENTING AI BIAS

One can think of machine intelligence as efficient, cold, and unbiased. In reality, AI can learn the biases of its programmers, which though likely unintentional, can have ramifications that could range from embarrassing to compromising.

The ability to distinguish, predict, and learn means that AI is able to think in a more abstract fashion than earlier forms of computing. To do this, AI needs to be trained with datasets and an orientation towards preferable outcomes. These training sets help an AI understand what destination it is supposed to aim for and which obstacles to avoid; the AI will figure out the optimal path on its own. Without enough training, an AI will have difficulty achieving its task, or will do so in a way that could lead to misinterpretations of data. The responsible policy manager needs to ensure that context important to the analysis is included; otherwise the AI runs the risk of interpreting data incorrectly.

[14] See IEEE Standards Association Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems: http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

The results can be problematic, and can reduce public and political perception that these tools are viable and useful. An AI can incorporate biases against ethnic groups due to systemic issues. As an illustrative example, there are currently a disproportionate number of indigenous persons in prison.[15] If an AI looking to flag potentially risky applications were to have access to that dataset without being taught how to avoid ethnic biases, it could flag applications made by on-reserve IP addresses as potentially risky simply because they originate on-reserve.

As AI applications are more widely dispersed throughout the economy, a number of these unintentional but grave biases have been uncovered. For example, according to a study by Carnegie Mellon University, women tend to be shown job ads for high-paying jobs less often than men as a result of search algorithms, likely due to the fact that women are disproportionately missing from these positions.[16]

Machines can't learn the concepts of equity or environmental stewardship without being taught that these goals – while maybe not the explicit goal of the system – are necessary trajectories to be taken into consideration. Understanding the risks posed by AI bias:

- How do we incorporate a variety of policy lenses into the design of an AI system? How do we teach policy managers how to recognize the potential pitfalls that may be found in data?
- What training do AI trainers need? What form of rules, guidelines, or best practices would be helpful in ensuring constructive and appropriate AI training?

PRIVACY

The privacy risks brought by AI are not fundamentally new; these are largely the same risks that exist with human agents, and often revolve around unacceptable collection or disclosure.

AI applications may pose less of a risk to individual privacy than human agents, as it is less likely to access and disclose personal information of a client without programmatic cause. An AI will not become curious about the personal details of a friend or favourite celebrity, for example. However, if an AI's access to information is unchecked to include all publically-available information, privacy concerns arise. Institutions must maintain strict control over what information the AI has access to.

- Under existing law, is the use of information by a machine treated differently than a human?
- What are the implications of AI for the privacy rights of individuals? Are existing legislative controls enough to manage these implications? What are some examples of hard-wired protections that can be implemented to protect privacy?

LANGUAGES

The *Treasury Board Policy on Official Languages* requires that institutions actively offer services to the public in both official languages, and of equal quality. An AI virtual agent is no exception to this policy; departments and agencies interested in deploying a virtual agent should ensure that it is capable of communicating fluently in both languages in spoken and textual form as appropriate.

[15] Office of the Correctional Investigator, Link: http://www.oci-bec.gc.ca/cnt/priorities-priorites/aboriginals-autochtones-eng.aspx

[16] 2015 study using 1,000 simulated persons, Link: http://www.cmu.edu/news/stories/archives/2015/july/online-ads-research.html

Virtual agent technology provides the opportunity to offer services in a multitude of other languages as well, depending on the ability to access necessary datasets for the AI to properly learn it. For example, an AI virtual agent can be taught a variety of indigenous languages, or some of the key international languages spoken in Canada. This would assist members of those communities in navigating application processes.

- Are there existing AI products on the market or in development capable of understanding and communicating in the French language to an aptitude equal to English?
- Should the Government of Canada use this technology to offer services in other languages? If so, to what degree? Could this create an expectation that these languages would be offered in other modes?
- Can virtual agents operating in French improve the performance of virtual agents operating in English, or do they have to develop separately by necessity?

LEGAL IMPLICATIONS

The law as it was conceived presupposes that humanity governs humanity. All of the institutions created by law are provided Ministers, deputy heads, and/or chief executive officers who are responsible and accountable for their conduct and for the execution of duties for which these institutions are mandated to perform. We are a people-driven government, but the introduction of a computer system that autonomously improves itself and modifies its instructions may blur the line between tool and agent under the law.

DELEGATION OF AUTHORITY

Typically, a minister's powers can be exercised by a delegate or another person who is in an appropriate capacity to exercise such powers. This is to ensure that responsibilities are managed with the reasonable assumption that one person could not possibly make every decision required in a department. The power of delegation is very common; it exists in program legislation, departmental enabling legislation, or broad authorities such as the *Financial Administration Act*. Based on subsection 24(2) of the *Interpretation Act*, It is also very common for persons to make decisions for and on behalf of a minister even when there is no explicit legislative power to delegate.

- Under current law, can an AI make a decision on behalf of a minister without human intervention, or must an AI be limited to a recommendation?
- Under current law, can an AI implement a decision taken at a higher level? Is it possible for several institutions to share one eligibility processing system for unrelated programs (e.g. as a cloud service), or do they have to be autonomous to ensure straightforward accountability?
- Is it possible for several institutions to share one eligibility processing system for unrelated, or do they have to be autonomous to ensure straightforward accountability?

ACCOUNTABILITY OF ADVICE

If a human service agent says something to a client that is inappropriate, or discloses another client's personal information accidentally, there is a continuum of accountability that extends from them upward through to their Minister. As responsible, cognitively-able human beings they are responsible for their actions, but their supervisor is also responsible for their training, and their supervisor for the design of training programs, and so on.

With an AI system providing service, the chain of accountability might look a bit different. To some extent the AI will say what it is scripted to, but it might learn conversational skills on its own with the practice it receives in the field. Similarly, if human policy analysts use AI simulations as the basis of their advice but the simulations turn out to be based on data incorrectly employed by the AI, who is responsible for the advice?

With this new form of service delivery:

- How does the accountability of advice change? Is accountability shared between the policy function and IT?
- How do we balance the potential of machine learning with the need to maintain accountability?

ADAPTABILITY TO CHANGES IN LAW AND POLICY

Imagine a scenario whereby an eligibility requirement for EI benefits has been interpreted by the Federal Court, who reverses or nuances how that provision has been interpreted by the EI Commission for the last 15 years. Under this scenario, not only would an AI have to be retrained, but its entire training set potentially adapted to meet this new interpretation.

While it is expected that the use of analytics produced by AI could help inform changes to policy, an AI would have to be retroactively adaptable to implement changes to rulesets.

- How will AI react to changes in eligibility criteria – whether policy, legislative, or judicial? Would the data sets that were used to inform/train the AI be relevant anymore or would a retraining period be required?

RESPONSES TO ACUTE DISTRESS

A client in direct interaction with a human agent could give open or subtextual indication that there is a threat of immediate harm to themselves or others. In such a circumstance a human would be guided by a mix of their training and their own moral compass, but machine intelligence would need some pre-programmed response. Consider a client that informs a virtual agent that they need help because they want to harm themselves.

- In the event of a distress situation, are there legal implications in case of an inappropriate response by an AI? Are there some circumstances where a machine should have access to more personal information to, for example, locate the individual at risk?

TECHNICAL ISSUES

AUDIBILITY AND CONTROL

Like a human, an AI needs to be trained to do its job. The training sets provided to it would establish its sense of right and wrong, which would be defined by the ethical standards described above. From a technical perspective, however, institutions need to know that the AI is following its training. With a human agent, conversations can be recorded, and there is an auditing mechanism to determine if programs are following their design appropriately. With an AI, audibility needs to be hardwired into the program.

Audits help configure an AI to ensuring that it is making appropriate decisions, so that it learns not to make the same errors again. They play the same role as fictional case training, except that they occur while the system is online. Audits in this case are akin to human quality assurance agents inspecting produced goods on an assembly line; auditors are essentially removing a decision from the queue, examining it for faults, and deciding whether to return the decision or change it.

A "Black Box?"

There is a segment of AI researchers that have recently described a problem colloquially known as the "black box problem." It states that while the algorithms that drive AI can be understood in when they are first written, as the system learns and grows, researchers aren't exactly sure of how the AI reaches outcomes. That is, while we can infer and predict what AI will come up with based on the data it has been given, we can't be 100% positive. This raises a legal question of how we can develop AI systems that we can describe in the context of a tribunal hearing or court case.

When a person denied a service appeals the decision to a tribunal or court, the government will need to explain the thought process that led to the decision. An AI eligibility determination system would be no exception; any such system would have to be able to generate an understandable rationale on every administrative decision that it made in case the decision leads to appeal in the future.

A related issue is control. If an AI system is capable of processing eligibility of thousands of applications in minutes, then it is potentially capable of considerable more damage if these decisions are made erroneously than the erroneous decisions of a single officer.

- How do we know that an AI is making a "right" decision? How does a policy manager know whether an AI eligibility processor is applying the same rigour of analysis as a trained, experienced human agent?
- How often should system audits be required? Should audits be transparent to the public?
- Can AI be employed to monitor the performance of another AI?

COGNITIVE INTUITION

A person arrives at the border and is subject to primary inspection by a Border Services Officer. In addition to the computerized tools available to her, the BSO takes a good look at the individual and uses her own experience and subconscious intuition to inform her decision whether to allow the person through, or recommend secondary inspection.

Intuition is a key element of human cognition. It is a composite of sensory inputs that our brain detects and analyses subconsciously. We use it to complement our conscious reasoning to make judgement calls.[17] AI does not differentiate between conscious and subconscious stimuli, and needs to understand these non-linguistic inputs well enough to include as evidence when making a decision.

[17] Likely. This is an evolving field. Scientific American Mind, "Intuition May Reveal Where Expertise Resides in the Brain," https://www.scientificamerican.com/article/intuition-may-reveal-where-expertise-resides-in-the-brain/

- In its current state, is machine vision advanced enough to detect non-linguistic communication? Is AI a capable complement for services where an in-person interview is required?

SYSTEM PERFORMANCE AND INTEROPERABILITY

The implementation of an AI system, either front-end or back-end, will require that a department has met certain technical prerequisites.

To start, an institution needs enough input data, and in formats that an AI can read and interpret. The implementation of an AI system, either front-end or back-end, will likely not happen in a vacuum. AI systems need to communicate with other systems that institutions use; for example, an AI eligibility processor may need to receive an application from the Web, communicate issues with human analysts, or email a client when a decision has been made.

The AI application requires a high degree of uptime and a low degree of latency. Communication with a client does not need to be instantaneous; interactions with human agents usually involve some lag or wait time as the agent researches information. Ideally the computer would be faster, as both speed and accuracy of responses build client trust. There will be higher levels of latency where internet access is limited to, for example, satellite access such as rural or remote communities.

- Are virtual agents effective for rural or remote clients with slow internet connections?
- What backup processes need to be in place to ensure that business works even when systems are down?
- How do we ensure that AI systems are secured from tampering?

WORKFORCE

There are widespread workforce implications to the automation brought by AI. In a post-AI Government of Canada, many tasks formerly undertaken by people would be replaced by machine intelligence. Private sector experiences in automation brought net reductions in workforce numbers. Automation can be perceived by staff as an existential threat, which can cause anxiety and stress that have negative impacts on mental health and productivity.[18]

In the short term, these technologies will mostly be used in an assistive fashion to improve productivity and reduce workload for staff. AI is a powerful tool that will supplement existing functions, allowing staff to focus on higher-value tasks. In the long term, as AI applications become more refined and systems upgraded in a way that expand their functionality, certain positions will likely be rendered obsolete as work changes. However, this has been the case throughout the history of labour, when mechanization and automation have changed the nature of work and allowed fewer people to produce more.

The automation of government functions via AI raises an ethical equation that is difficult to solve: on one hand, the potential for a reduction in positions in the medium-term; on the other, a potential to improve the rapidity and efficiency of service.

Numbers are only one part of the story. The shift towards AI automation brings with it a challenge to the Government of Canada to ensure that we have highly-skilled persons capable of understanding and working with

[18] The Economist, "Automation and Anxiety," June 2016. Link: http://www.economist.com/news/special-report/21700758-will-smarter-machines-cause-mass-unemployment-automation-and-anxiety

this technology. As virtual agents improve and the need for front-line service agents diminishes, there are opportunities for staff to have their skills upgraded to deal with complex cases.

- What are the skills that employees will need in post-AI government? How do we assist that transformation in a way that ensures that employees feel as supported as possible?

COMMUNICATIONS

Science fiction and popular media have depicted this technology in both positive and negative lights, and it is increasingly reported in mainstream and specialized news outlets. Despite its profusion in today's economy, AI has often been referred to by popular culture negatively.[19] To avoid opposition to this technology, a widespread deployment of AI will require careful and transparent communications with the public to be successful.

This is especially true with a virtual agent, which can act as a direct representative between clients and the federal government. Thought would have to be placed on the language it would use, what persona if any it would employ, and the response institutions should pursue if it acted in a way that was inappropriate.

- How should federal use of AI be communicated to the public?
- Should we report on AI systems' performance?
- What are the design considerations that should be included in the deployment of a virtual agent program?

CONCLUSION

Artificial intelligence is neither science fiction, nor a technology restricted to laboratories. Chances are that your life is already affected by AI. Developers large and small are researching and iterating at a feverish pace in this space, ready to offer products that can directly improve the quality and timeliness of public services. There are strong potential benefits to this technology that can bring real value to all of the Government of Canada's clients in the short term, but there are also implications that need to be addressed quickly and methodically before this it is exploited to its fullest potential. However, this is a fundamentally disruptive technology, and there will be some questions that will only be answered once it is deployed and in use.

[19] (Canadian source pending, but have British Science Association survey as evidence of this trend: https://www.britishscienceassociation.org/news/rise-of-artificial-intelligence-is-a-threat-to-humanity)

Pages 105 to / à 116

are withheld pursuant to section

sont retenues en vertu de l'article

69(1)(a)

of the Access to Information Act

de la Loi sur l'accès à l'information